Deposited as pre-print on *bioRxiv*

Metagenomic profiling and transfer dynamics of antibiotic resistance determinants in a full-scale granular sludge wastewater treatment plant

David Calderón-Franco^a, Roel Sarelse^a, Stella Christou^a, Mario Pronk^{a,b}, Mark C.M. van Loosdrecht^a,
Thomas Abeel^{c,d,§}, David G. Weissbrodt^{a, §, *}

^aDepartment of Biotechnology, Delft University of Technology, Delft, The Netherlands

^bRoyal HaskoningDHV, Amersfoort, The Netherlands

^cDelft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands

^dInfectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, USA

[§] Shared senior authorship

*Correspondence: David G. Weissbrodt, Assistant Professor, Weissbrodt Group for Environmental
Life Science Engineering, Environmental Biotechnology Section, Department of Biotechnology, Delft
University of Technology, van der Maasweg 9, 2629 HZ Delft, The Netherlands; E-mail:
d.g.weissbrodt@tudelft.nl; Tel: +31 15 27 81169

Abstract

In the One Health context, wastewater treatment plants (WWTPs) are central to safeguard water resources. Nonetheless, many questions remain about their effectiveness to prevent the dissemination of antimicrobial resistance (AMR). Most surveillance studies monitor the levels and removal of selected antibiotic resistance genes (ARGs) and mobile genetic elements (MGEs) in intracellular DNA (iDNA) extracted from WWTP influents and effluents. The role of extracellular free DNA (exDNA) in wastewater is mostly overlooked. In this study, we analyzed the transfer of ARGs and MGEs in a full-scale Nereda® reactor removing nutrients with aerobic granular sludge. We tracked the composition and fate of the iDNA and exDNA pools of influent, sludge, and effluent samples. Metagenomics was used to profile the microbiome, resistome, and mobilome signatures of iDNA and exDNA extracts. Selected ARGs and MGEs were analyzed by qPCR. From 2,840 ARGs identified, the genes *arr-3* (2%), *tetC* (1.6%), *sul1* (1.5%), *oqxB* (1.2%), and *aph(3'')-Ib* (1.2%) were the most abundant among all sampling points and bioaggregates. *Pseudomonas*, *Acinetobacter*, *Aeromonas*, *Acidovorax*, *Rhodoferax*, and *Streptomyces* populations were the main hosts of ARGs in the sludge. In the effluent, 478 resistance determinants were detected, of which 89% from exDNA potentially released by cell lysis during aeration in the reactor. MGEs and multiple ARGs were co-localized on the same extracellular genetic contigs. These can pose a risk for AMR dissemination by transformation into microorganisms of receiving water bodies. Total intracellular ARGs decreased 3-42% as a result of wastewater treatment. However, the *ermB* and *sul1* genes increased by 2 and 1 log gene copies mL⁻¹, respectively, in exDNA from influent to effluent. The exDNA fractions need to be considered in AMR surveillance, risk assessment, and mitigation.

Keywords: Aerobic granular sludge; Free-floating extracellular DNA; Intracellular DNA; Antibiotic resistance genes; Mobile Genetic Elements; qPCR; Metagenomics

45 Graphical abstract

46

Highlights

- A DNA database from an AGS reactor was constructed to study the system resistome, mobilome, and microbiome.
- The genera *Pseudomonas* and *Rhodoferax* were the predominant ARG carriers in the system.
- MGEs and ARGs often co-localize on contigs recovered from the exDNA of the effluent.
- AGS plants are efficient at reducing ARB.
- The exDNA is an underestimated DNA fraction containing ARGs in the effluent.

List of abbreviations

<i>aac(3)-Ib</i>	Aminoglycoside 3'-N-acetyltransferase resistance gene
AMR	Antimicrobial resistance
<i>aadA6</i>	Aminoglycoside (3'') adenytransferase resistance gene
<i>aadA11</i>	Aminoglycoside (3'') (9) adenytransferase resistance gene
AGS	Aerobic granular sludge
<i>aph(3'')-Ib</i>	Aminoglycoside 3'-phosphotransferase resistance gene
<i>aph(6)-Id</i>	Aminoglycoside O-phosphotransferase resistance gene
ARB	Antibiotic-resistant bacteria
ARG	Antibiotic resistance gene
<i>arr-3</i>	Plasmid-encoded ribosyltransferase resistance gene
BGC	Biosynthetic gene clusters
COG	Clusters of Orthologous Genes
exARG	Free-floating extracellular antibiotic resistance gene
exDNA	Free-floating extracellular DNA
HGT	Horizontal gene transfer
iARG	Intracellular antibiotic resistance gene
iDNA	Intracellular DNA
MAG	Metagenome-assembled genome
MGE	Mobile genetic element
<i>oqxB</i>	Fluoroquinolone efflux pump membrane transporter gene
qPCR	Quantitative polymerase chain reaction
SBR	Sequential batch reactor
<i>sul1</i>	Sulfonamide resistant dihydropteroate synthase of Gram- bacteria, linked to class 1 integrons
<i>tetC</i>	Tetracycline efflux pump gene
WWT	Wastewater treatment
WWTP	Wastewater treatment plant

1. Introduction

Drug-resistant diseases currently cause at least 700,000 deaths globally per year. This number could increase to 10 million y^{-1} by 2050 across all income regions, under the most alarming scenario, which is that no action is taken to contain antimicrobial resistance (AMR) (WHO & UN, 2019). The proliferation of antibiotic-resistant bacteria (ARB) directly correlates with the widespread use of corresponding antibiotics (Davies and Davies, 2010). Medically problematic pathogens that acquired multidrug resistance through misuse of antibiotics are bacteria responsible for tuberculosis, acute respiratory infections, sexually transmitted infections, and bacillary dysentery (Ukuhor, 2021).

Within the One Health context, wastewater treatment plants (WWTPs) should form a barrier between sewage that transports high loads of antibiotics and antibiotic resistance genes (ARGs) emitted by anthropogenic metabolisms, and aquatic ecosystems (Bürgmann et al., 2018). Microorganisms in WWTPs biological processes are often considered reservoirs of ARGs, while also supposed to contribute in mitigating AMR dissemination by degrading antibiotics and AMR determinants. However, current WWTP designs are not optimized to this end.

Currently, AMR surveillance in wastewater is primarily conducted with molecular biology measurements that target the intracellular DNA (iDNA) of ARB. By examining the influent and effluent of 62 Dutch WWTPs by qPCR, Pallares-Vega et al. (2019) observed not only a reduction in the abundance of ARGs, but also an increase in the relative abundance of resistance plasmids of the incompatibility group 1 (IncP-1). Guo et al. (2017) used metagenomics to describe microbiome, mobilome, and resistome patterns from the iDNA pool of a Chinese WWTP, revealing that *Clostridium* and *Nitrosomonas* can carry ARGs during WWT.

Besides identifying at high resolution the hosts of ARGs in microbial communities of activated sludge, it remains primordial to elucidate the mechanisms and mobile genetic elements (MGEs) that transfer AMR determinants in these populations.

In addition to iDNA, extracellular free DNA (exDNA) contains a high proportion of MGEs (Calderón-Franco et al., 2020). Different genetic structures and architectures (plasmids, transposons, insertion sequences, and integrases, among others) transfer ARGs between bacteria, but many questions remain. The analysis of both intracellular ARGs (iARGs) and extracellular ARGs (exARGs) combined with mobilome co-localization analysis has rarely been performed in complex environmental samples such as wastewater biomasses. Such co-localization analysis involves the characterization of the

resistome fraction in genomic proximity to horizontal gene transfer (HGT) mediators such as plasmids and other mobile genetic elements (Slizovskiy et al., 2020).

Apart from the presence of exDNA in wastewater, not much is known about the actual transfer of AMR in the microbiome of WWTPs and the underlying effects of biofilms. Dense microbial aggregations can promote the horizontal transfer of antibiotic resistance genes (ARGs) in aquatic environments (Abe et al., 2020; Madsen et al., 2012). Full-scale Nereda® plants that use aerobic granular sludge (AGS) for an integral removal of nutrients (Pronk et al., 2015) harbor a hybrid sludge composed of granules (0.2-3 mm) that have similar properties as biofilms. Therefore, AGS SBRs form interesting microbial ecosystems to study the fate of AMR. Metagenomic analysis during granulation at pilot scale showed that ARGs enriched in both iDNA and exDNA fractions of AGS during the granule development stage, and that integrons played an essential role in carrying exARGs (Li et al., 2020).

In this study, we analyzed the transfer dynamics of ARGs and MGEs in a full-scale AGS Nereda® plant. Metagenomic and qPCR analyses were performed on the iDNA and exDNA pools of samples collected from the influent, different sludge fractions and the effluent. The sludge was sampled over the different SBR phases (end of anaerobic feeding, end of aeration) and sifted for different sized bioaggregates (flocs, small and big granules), to track transfer phenomena. The findings highlight the fate of the AMR determinants in a full-scale AGS WWTP and the importance of considering both iDNA and exDNA pools in AMR dissemination studies.

2. Materials and methods

2.1. Sampling of a full-scale AGS WWTP

The sampling campaign performed on a Nereda® plant located in Utrecht (52.11215, 5.10813), The Netherlands. Sampling was performed over 3 days (Monday, Wednesday, and Friday) within the same week, under dry weather conditions without significant variations in hydraulic retention time (20 hours). Since the sludge retention time was relatively long (30 days in average; >40 days for large granules and <10 days for small granules and flocs) in this installation, the 3 sampling days were considered as biological replicates.

Biological samples were collected every day from pre-settled influent wastewater, from the mixed liquor at the beginning (*i.e.*, after anaerobic feeding of the wastewater) and end of the aeration phase, and from the effluent of the SBR (**Figure 1a**). Total volumes of 1 L of influent and 1 L of effluent were collected as a 24-h flow-proportional composite samples. For the other process points, the sludge was sampled from the top of the tank as grab samples of 1 L with a pole container during the daily operations of the SBR. All samples were transported in a cooled container and processed within a time frame of less than 4 h prior to DNA extraction.

The granular sludge samples were sieved in 3 different fractions of bioaggregates (flocs <0.2 mm, small granules of 0.2-1 mm, and large granules >1 mm) according to Ali et al. (2019). The water at the outlet of the sieving unit was used for exDNA isolation. The fractions of exDNA were isolated in a time-frame of less than 4 h after sampling. The sieved fractions of granular sludge were stored at -20°C in a time-frame of less than 4 h after pending isolations of iDNA.

In total, 12 analytes were obtained and sequenced by metagenomics per sampling day, *i.e.*, 36 analytes for the whole campaign. **Table 1** summarizes the set of samples and analytes.

Table 1. Samples collected from the influent, sludge, and effluent of the SBR during each of the 3 sampling days at WWTP Utrecht. Granular sludge samples were sieved in 3 different fractions of flocs, small granules, and large granules. The pools of intracellular DNA (iDNA) and free-floating extracellular DNA (exDNA) extracted from each sample as analytes for metagenomics are indicated. Two out of the three biological replicates were sequenced for metagenomics analysis, indicated as “-D1” and “-D2” for samples taken on Monday and Friday, respectively.

SBR operation stage sampled	Number of successive days sampled (<i>i.e.</i> ,	Type of biological sample and bioaggregate size	DNA pool isolated from	Analyte metagenomic samples Identifier	BioSample IDs
-----------------------------	---	---	------------------------	--	---------------

	biological replicates)		the sample		
Influent	3	Global sample	exDNA	INF-exDNA-D1 INF-exDNA-D2	SAMN23455551 SAMN23455563
			iDNA	INF-iDNA-D1 INF-iDNA-D2	SAMN23455550 SAMN23455562
Start of aeration	3	Global sample	exDNA	BA-exDNA-D1 BA-exDNA-D2	SAMN23455555 SAMN23455567
		Flocs	iDNA	BA-FL-iDNA-D1 BA-FL-iDNA-D2	SAMN23455554 SAMN23455566
		Small granules	iDNA	BA-SG-iDNA-D1 BA-SG-iDNA-D2	SAMN23455553 SAMN23455565
		Large granules	iDNA	BA-LG-iDNA-D1 BA-LG-iDNA-D2	SAMN23455552 SAMN23455564
End of aeration	3	Global sample	exDNA	EA-exDNA-D1 EA-exDNA-D2	SAMN23455559 SAMN23455571
		Flocs	iDNA	EA-FL-iDNA-D1 EA-FL-iDNA-D2	SAMN23455558 SAMN23455570
		Small granules	iDNA	EA-SG-iDNA-D1 EA-SG-iDNA-D2	SAMN23455557 SAMN23455569
		Large granules	iDNA	EA-LG-iDNA-D1 EA-LG-iDNA-D2	SAMN23455556 SAMN23455568
Effluent	3	Global sample	exDNA	EFF-exDNA-D1 EFF-exDNA-D2	SAMN23455561 SAMN23455573
			iDNA	EFF-iDNA-D1 EFF-iDNA-D2	SAMN23455560 SAMN23455572

2.2. Extractions of exDNA and iDNA analytes

The pools of iDNA and exDNA were differentially isolated for all samples, as recently described by Calderón-Franco et al. (2020).

Filtrations of influents and effluent samples. The total volumes of 1 L of influent and effluent were filtered through 0.22 µm polyvinylidene fluoride (PVDF) membranes (Pall Corporation, USA).

The filtrates were used to isolate the exDNA. The biomasses retained on membranes were used for iDNA extractions.

Centrifugations and homogenizations of sludge fractions. The sieved sludge fractions were centrifuged at 4,000 ×g for 10 min before storing at -20°C. The biomass pellets were used for iDNA extraction. The pellets were first treated with DNase I to remove remaining exDNA fragments, and homogenized with a Potter-Elvehjem pestle (Wheaton, USA).

Isolations of exDNA. The water fractions remaining after sieving were used for exDNA isolation. After filtration through 0.22 µm PVDF membranes, exDNA was loaded for ion-exchange chromatography

on a 1-mL, positively charged, diethylaminoethyl cellulose (DEAE) column with 2- μ m pore size (BIA Separations, Slovenia). Manufacturer's instructions were followed to condition the column and to process the chromatographic operations. The following buffers and solutions were used to equilibrate, elute, regenerate, clean, and store the DEAE column: (i) equilibration buffer (50 mmol L⁻¹ Tris, 10 mmol L⁻¹ EDTA at pH 7.2); (ii) elution buffer (50 mmol L⁻¹ Tris, 10 mmol L⁻¹ EDTA, 1.5 mol L⁻¹ NaCl at pH 7.2); (iii) regeneration buffer (50 mmol L⁻¹ Tris, 10 mmol L⁻¹ EDTA, 2 mol L⁻¹ NaCl at pH 7.2); (iv) cleaning solution (1 mol L⁻¹ NaOH and 2 mol L⁻¹ NaCl); (v) storage solution (20% ethanol in ultrapure water). The column was equilibrated using the equilibration buffer at a flowrate of 0.6 mL min⁻¹ to maintain the pressure below the maximum limit of 1.8 MPa. After equilibration, the entire volume (1 L) of filtrates containing the exDNA was loaded in the DEAE column using an HPLC pump (Shimadzu Corporation, Japan) at the same flowrate of 0.6 mL min⁻¹. After retention, exDNA was eluted at a flow rate of 1 mL min⁻¹ using the elution buffer and tracked over time with an HPLC UV-Vis photodiode array detector (Waters Corporation, USA) recording the UV-light absorption of DNA at 260 nm. The recovered raw exDNA was precipitated from the eluate with ethanol (Moore and Dowhan, 2002). The residual proteins bound to exDNA were digested by 2-h incubation with 0.85 g L⁻¹ proteinase K (Sigma-Aldrich, UK) at room temperature. This enzymatic reaction was stopped at 50°C for 10 min in a thermoblock (ThermoMixer, Eppendorf). The protein-digested exDNA was finally purified using GeneJET NGS Cleanup Kit (Thermo Scientific, USA). The purified exDNA isolates were stored at -20°C pending analysis.

Extractions of iDNA. The membranes containing the biomasses filtrated from the influent and effluent samples were directly used for iDNA extraction using PowerWater DNA extraction kit (QIAGEN, The Netherlands) following manufacturer's instructions. For the bioaggregate samples, a mass of 0.25 g of wet weight of homogenized biomass was used to extract iDNA. IDNA extractions were performed using the DNeasy PowerSoil DNA extraction kit (QIAGEN, The Netherlands) following manufacturer's instructions.

2.3. Library preparation, sequencing, quality control, assembly, and binning

Selection of analytes. From the set of 36 analytes of exDNA and iDNA obtained over the 3 days, the analytes from the Monday and Friday (*i.e.*, 24 analytes) were selected as duplicates (economic optimum) for metagenomics. These exDNA and iDNA analytes were sent to DNASense (Aalborg, Denmark) for library preparation and sequencing.

Preparation of libraries. The DNA was quantified using Qubit (Thermo Fisher Scientific, USA) and fragmented to approximately 550 bp using a Covaris M220 with microTUBE AFA Fiber screw tubes and the settings: Duty Factor 10%, peak/displayed power 75 W, cycles/burst 200, duration 40 s and temperature 20°C. The fragmented DNA was used for metagenome preparation using the NEB Next Ultra II DNA library preparation kit.

Sequencing of libraries. Libraries were sequenced using the HiSeq sequencer (Illumina, USA) as 2x150 bp paired-end reads.

Quality control of sequence reads. After sequencing, a dataset containing 48 paired-end read samples with an average of 16 million reads per sample was obtained. Metagenomics workflow is summarized in **Figure S1** in supplementary information. The minimum and maximum numbers of quality-filtered, non-duplicated sequencing reads of 150 bp ranged from 12 to 18 million (**Figure S2** in supplementary information). The quality of the Illumina reads was assessed using FastQC version 0.11.9 with default parameters (Andrews, 2010). Low-quality paired-ends reads were trimmed and filtered by Trimmomatic version 0.39 on paired-end mode (Bolger et al., 2014).

Assembly of sequence reads. The trimmed reads were assembled into contigs using metaSPAdes version 3.14.1 (Nurk et al., 2017) on meta mode on default parameters.

Binning of DNA contigs. Contigs resulting from the sequencing of only the iDNA pools of bioaggregates were binned with MetaBAT version 2.2.15 (Kang et al., 2019) to reconstruct metagenome-assembled genomes (MAGs) on default parameters. The MAGs were used to analyze secondary metabolite biosynthesis gene clusters.

2.4. Generation of the taxonomic database of aerated granular sludge

A database of sequences from MAGs, contigs, and reads specific to the microbial environment of AGS was built to accurately profile the microbiome. The Kraken 2.0 standard database of 9.1 Gbp of genomic sequences (Wood et al., 2019) was used as a basis. However, because complete genomes of the organisms found in AGS systems were often unavailable in the Kraken 2.0 database, additional genetic fragments specific to some taxa were added to the database. As an example, abundant genera in AGS like *Tetrasphaera*, *Trichococcus* and “*Candidatus Accumulibacter*” were not or poorly

annotated in the standard database used in Kraken2.0. In total, 94,005 sequences of 2,223 unique taxa were added to the Kraken 2.0 standard database.

The taxa of the AGS samples were classified from both short reads and contig sequences by combining the MetaPhlAn3.0 (Truong et al., 2015), MG-RAST (Meyer et al., 2018) and Kraken2.0 (Wood et al., 2019) computational tools for phylogenetic analysis of metagenomics data with their corresponding databases, as well as BLASTn to align contigs of >1500 bp against the MiDAS database of 16S rRNA gene sequences with a cut-off e-value $<10^{-5}$ and sequence identity >97% [<https://www.midasfieldguide.org/>] (Nierychlo et al., 2020). A literature study was conducted in which taxa were added to the database if theoretically present above 1% relative abundance but not present in the database. Kraken2.0 takes all the sequences added in a selected database and finds k-mers (short genomic substrings specific for certain taxa) for further taxonomic classification.

The reads (full and partial genomes) of the predominant taxa were added to the adapted taxonomic database of AGS, when these lineages were detected above 1% of relative abundance from the metagenomics datasets using one of the previously described classification tool but not present in the Kraken2.0 database. This was performed on taxonomic levels from phylum to species.

2.5. Microbiome profiling of iDNA and exDNA pools obtained from the AGS process

Kraken2.0 with short-reads as input was selected to profile the microbiome. Classification with Kraken2.0 was performed on pair-end mode on the quality-controlled short reads, using the newly constructed database.

The taxonomic classification was also performed on contigs >500 bp that were identified to contain ARGs (see § 2.6 hereafter), in order to determine potential ARGs hosts.

The taxonomic classification outcomes from Kraken2.0 were converted into abundance tables using the Pavian visualization tool (Breitwieser and Salzberg, 2020) to explore metagenomics classification datasets. Heatmaps were generated with the R package “pheatmap” (Kolde, 2019).

2.6. Resistome and mobilome profiling of iDNA and exDNA obtained from the AGS process

ARGs were annotated by aligning the assembled contigs >500 bp to the ResFinder 4.0 resistance gene database using the BLASTn nucleotide alignment tool with a cut-off E-value of $<10^{-5}$ and sequence identity above 90% (Bortolaia et al., 2020).

MGEs were classified on the same set of contigs >500 bp using BLASTn by aligning them with the same cut-off and sequence identity setpoints to the following databases depending to the types of MGEs. Bacterial insertion sequences were identified by aligning against the ISfinder database (Siguier et al., 2006). Integrons, integrases, and gene cassettes were identified using the INTEGRALL database (Moura et al., 2009). Bacterial integrative and conjugative elements were identified using the ICEBERG database (Liu et al., 2019).

For all queries, the ARG or MGE identified with the best score (*i.e.*, equal to the sequence identity multiplied by the coverage factor) was selected to annotate the query.

BLASTn was performed with different databases (ResFinder 4.0 for ARGs and ISfinder, INTEGRALL and ICEBERG for MGEs) on the same set of contigs to identify where ARGs and MGEs co-localized. BLASTn was performed with a cut-off E-value of $<10^{-5}$ and sequence identity above 90%. Contigs >500 bp that simultaneously contained hits from the ResFinder 4.0 database and at least one of the different MGE databases were considered to have co-localized.

The contigs that contained both ARGs and MGEs were used as input queries against the NCBI plasmid database in order to know if such contig belonged to a plasmid. A contig was identified as part of a potential plasmid if exclusively aligning with plasmids in the entire NCBI plasmid database with sequence identities >90%.

2.7. Functional genetic analysis of iDNA and exDNA pools obtained from the sludge

Prokka version 1.14.5 was used to annotate the assembled contigs >500 bp, with the default databases and parameters on metagenomic mode (Seemann, 2014). K-numbers were assigned to all predicted coding sequences (CDSs) using GhostKoala, after which the KEGG database was used for analyse the functional and AMR pathways (Kanehisa et al., 2002; Seemann, 2014). The antibiotics and secondary metabolites analysis shell (antiSMASH, v5.0) tool was used to identify, annotate, and analyze the secondary metabolite biosynthesis gene clusters, such as involved on antibiotics production, in MAGs binned from the iDNA bioaggregate samples (Blin et al., 2021).

2.8. Multidimensional scaling analysis to cluster microbiome and resistome datasets

To identify patterns between the microbiome and resistome, different types of ordination and statistical numerical methods were performed using the R software. The input data was the output of the taxonomic classification from Kraken2.0 with the adapted database.

Dimension reduction is a projection-based method that transforms the data by projecting it onto a set of perpendicular axes. Before dimension reduction, the metagenomic datasets were normalized by dividing the measurements by the corresponding sample size factor using software package "edgeR" (Robinson et al., 2009), and a variance stabilization transformation (VST) was performed using "DESeq2" (Love et al., 2014), respectively. The "vegan" package (Dixon, 2003) transformed the stabilized matrices into distance matrices. The Bray-Curtis distance metric was identified as the optimal distance metric for all dimension reduction methods.

The dimensions of the metagenomics datasets were reduced using Non-metric Multidimensional scaling (NMDS), Principal Coordinates Analysis (PCoA), and t-distributed stochastic neighbor embedding (t-SNE). These methods were applied to the taxonomic data at phylum and genus levels, and to the AMR group data. For all methods, scree-plots, stress-plot, and Shepard-plots were used to visualize the data accurately in two dimensions. t-SNE was applied using the "tsne" package, and a maximum of 2,000 iterations and a perplexity parameter of 5 as input (van der Maaten and Hinton, 2008). PCoA, t-SNE, and NMDS were performed on different taxonomic levels and on the resistome data (**Figure S3**). All dimension reductions yielded similar clustering patterns, giving equivalent results for each sampling point and sample type. PCoA was therefore chosen as representative analysis to characterize the similarities and differences between samples.

2.9. Quantitative PCR of selected ARGs

To evaluate the WWTP performance in terms of removing ARGs and MGEs, to examine their presence during the different process steps, and to track their transfer between iDNA and exDNA fractions, a molecular analysis by quantitative polymerase chain reaction (qPCR) was applied to detect a selected panel of genes (**Table 2**) from the three biological replicates (Monday, Wednesday, and Friday) of the samples listed in **Table 1**.

All iDNA and exDNA analytes isolated from the three-biological sample from the influent, beginning and end of aeration, and effluent of the SBR were used for qPCR. Each analyte was measured in technical duplicates by qPCR.

The *16s rRNA* and *rpoB* reference genes were quantified for the normalization of gene copies to the concentration of bacteria in the iDNA samples.

Based on antibiotics consumption data in the Netherlands (RIVM and SWAB, 2020), ARGs were targeted from the antibiotic groups of beta-lactams (*bla_{CTXM}*), macrolides (*ermB*), fluoroquinolones (*qnrS*), sulfonamides (*sul1/sul2*), and tetracyclines (*tetO*). The *intI1* gene encoding the class I integron-integrase was quantified. These integrase class I cassettes are related to ARG mobility, acquisition, and exchange between microorganisms (Ma et al., 2017).

Primers, thermal cycler conditions, and gBlocks gene fragments used for standards generation and quantification during qPCR are given in **Tables S1, S2, and S3**, respectively.

Table 2. ARGs and MGE selected for qPCR with their description

Category	Gene	Description
Bacteria normalization	<i>16S rRNA</i>	RNA component of the 30S small subunit of prokaryotic ribosome
	<i>rpoB</i>	Bacterial RNA polymerase subunit β
ARGs	<i>bla_{CTXM}</i>	Bla-cefotaxime-hydrolyzing β -lactamase
	<i>ermB</i>	Erythromycin resistance via methylation of 23S rRNA
	<i>qnrS</i>	Quinolone resistance protein
	<i>sul1</i>	Sulfonamide resistant dihydropteroate synthase
	<i>sul2</i>	Sulfonamide resistant dihydropteroate synthase
	<i>tetO</i>	Tetracycline resistance protein TetO
MGE	<i>Int1</i>	Class 1 integron integrase

2.10. Analysis of qPCR results

R version 3.6.3 (R Foundation for Statistical Computing., 2018) was used for visualization and statistical analysis of qPCR data.

A non-parametric Wilcoxon test was performed to determine the significance of the differences detected in the concentrations of genes ($\log_{10}$ gene copies VSS⁻¹ for iDNA or mL⁻¹ for exDNA) between the iDNA and exDNA fractions, and in the absolute and relative abundances of these genes in the influent and effluent (*i.e.*, indicating a removal of the genes due to treatment).

388

389 The same Wilcoxon test was used to indicate the significant difference between *rpoB* and *16S rRNA*
390 gene copies for their use as normalization factors in the qPCR results.

391

392 Statistical significance was established at the 95% confidence level ($p < 0.05$). The p-values for the
393 Wilcoxon test are given in Supplementary Material (**Table S4-S5**).

3. Results and discussion

The iDNA and exDNA pools were isolated from samples collected over 3 sampling days (Monday, Wednesday, Friday) out of the influent, mixed liquor (flocs, small and large granules), and effluent of a full-scale AGS SBR of WWTP Utrecht. The Monday and Friday iDNA and exDNA samples were metagenomically sequenced as biological duplicates to analyze their microbiome, resistome, and mobilome. The construction of the taxonomic database of AGS environment from the metagenomes was an essential step and first result to this end. Numerical analyses were used to map and identify patterns in these metagenomic data, with a focus on the ARG and MGE in microbiome of the WWTP. All samples from the 3 days were subsequently measured by qPCR to track selected ARGs and MGEs in the different samples.

3.1. Taxonomic composition of exDNA mainly linked to the granular sludge fraction

After taxonomic classification at genus level of the DNA contigs assembled from the metagenomic sequencing reads, principal coordinates analysis (PCoA) and other multidimensional scaling methods (**Figure S3** in supplementary material) were efficient to observe the clustering effects in data related to the specific sample DNA fraction (**Figure 1b**).

The influent samples showed a large difference in composition for both the iDNA and exDNA pools. Microbial compositions of sewage can vary over an active week, based on the variety of emission sources and streams in a wastewater catchment area and environmental factors (Pallares-Vega et al., 2021).

In the effluent, a lower variability between the two samples was observed for the iDNA fraction. The taxonomic composition of the iDNA of the effluent highly resembled the composition of the iDNA of activated sludge floc fraction in the treatment reactor. On the other hand, the phylogenetic signatures of the exDNA fraction of the effluent showed a larger variation between the samples.

In the wastewater treatment plant microbial cells can lyse, releasing iDNA as exDNA. The exDNA fractions might not persist as “free-floating” during the whole process. It can adhere to particles, get degraded, or taken up via natural transformation in competent cells (Dong et al., 2019; Zhou et al., 2019).

From the different size fractions of bioaggregates in the sludge, the microbial community compositions of the small (blue in **Figure 1b**) and large (red) granules were similar, while the flocs (orange) clustered separately. This is similar as the observations by Ali et al (2019) that the floc fraction is more diverse and contains a large number of bacteria originating from the influent wastewater. Microbial niches establish along gradients of substrates and other dissolved components like oxygen and nitrogen oxides inside bioaggregates (Winkler et al., 2018).

The taxonomic affiliations of the exDNA sequences of the mixed liquor sampled at the end of aeration were similar to those of the iDNA extracted from small and large granules, rather than the flocs. This suggests that exDNA was released from the granules by cell lysis during aeration. Sengar et al. (2018) and Toh et al. (2003) have shown that dead biomass releases its genetic material into the extracellular environment during the AGS process. In activated sludge, Yuan et al. (2019) have identified that the taxa reflected by the iDNA and exDNA fractions are similar, suggesting a correlation between changes in the microbial composition of the activated sludge and in exDNA. Using the same exDNA extraction methodology, we have shown in a previous study that exDNA present in activated sludge mainly originates from decaying microbial populations rather than from transport from the influent (Calderón-Franco et al., 2020).

3.2. AGS microbiome and involvement in AMR dissemination

The patterns of relative abundances of genera in the different stages of the AGS process and in the iDNA and exDNA fractions were examined by constructing a heatmap (**Figure 1c**) for the clusters observed by PCoA.

Using the standard Kraken2.0 database (NCBI based), 2,548 genera were identified from the whole set of iDNA and exDNA analytes: 57 genera were more abundant than 1% of sequencing read counts, and 15 of those were predominant above 3% in at least one of the samples (**Figure S4, * symbols**). The average annotation rate of the reads was $24.7 \pm 3.2\%$, *i.e.*, leaving three quarters of the taxonomic information of the reads as to uncover.

To increase the reads annotation rate, a new database of genetic sequences of the AGS microbial environment was constructed. With this database, the average annotation rate was $27.1 \pm 3.4\%$; 2,679 genera were identified: 55 genera were above 1% relative abundance and 21 were above 3% in at least one of the analytes (**Figure S4, ◇**). Although this new annotation rate was only 3% higher than

the standard database, important AGS genera were now identified. Metagenomic studies can be hindered by low annotation rates as a consequence of the high fraction of natural microbes that have not been included in the databases. The current development of databases of high-quality MAGs and properly curated databases will help enhance annotation rates in the future (Singleton et al., 2021).

From **Figure 1b**, the exDNA taxonomic compositions at the beginning of aeration resembled the exDNA from influent while at the end of aeration from both days closely resembled the iDNA fractions from small and large granules. Flocs displayed a unique profile. While the phylogenic distribution of the effluent exDNA resembled the granule phylogeny, the effluent iDNA resembled the flocs (from both beginning and end of aeration). During the fill/draw operation of the SBR, flocs leave the tank with the selection spill, therefore reducing their retention time in the process (Guo et al., 2020).

The diversity of abundant populations in the AGS detected by metagenomics is listed in **Figure 1c**. Several of these populations are involved in the conversions of C-N-P nutrients in the microbial ecosystem of AGS (Weissbrodt et al., 2014; Winkler et al., 2018), like “*Ca. Accumulibacter*”, *Tetrasphaera*, *Dechloromonas* (C, P), “*Ca. Competibacter*”, *Zoogloea*, *Xanthomonas*, *Rhodoferax*, *Pseudomonas*, *Acidovorax*, *Comamonas*, *Rhizobium* (C, N), *Nitrospira* (N), and *Acinetobacter* (C) among others. Some relate to filamentous bulking phenomena like “*Ca. Microthrix*”, *Bulkholderia*, *Nocardioides*, *Kouleothrix*. Some thrive on metabolites and lysis products from other organisms like *Flavobacterium*, *Bacteroides*, and *Hydrogenophaga* among many others. The metabolic functions of other organisms in the AGS remains to be uncovered. Less beneficial organisms are *Arcobacter* and *Aeromonas* that are known pathogens. *Arcobacter* crosses WWTPs from influent to effluent without settling properly (Kristensen et al., 2020). The presence in the influent of exDNA affiliating with “*Ca. Accumulibacter*” can relate to the recirculation line coming from the sludge thickener.

Among the different populations identified from the AGS (**Figure 1c**), several are known to carry AMR-related genes. *Acidovorax* ($3.5 \pm 0.6\%$) is a known carrier of beta-lactamase resistance plasmids in activated sludge samples (Zhang et al., 2011). *Rhodoferax* ($3.5 \pm 0.5\%$) has been identified to be of high relative abundance in wastewater effluent, while often identified as a carrier of resistance genes (Zhou et al., 2020). *Nitrospirae* ($4.1\% \pm 1.1\%$) may play a role in AMR dissemination because it is a known host of iARGs and exARGs in WWTPs (Zhou et al., 2019). Guo et al. (2017) used metagenomics to identify “*Candidatus Accumulibacter*” ($2.1 \pm 0.5\%$) as a possible carrier of resistance genes over the WWT operation. *Aeromonas* was relatively abundant in all our samples ($1.88 \pm 0.84\%$). Notably, the three most critical antibiotic-resistant pathogens as designated by the World Health Organization are

Acinetobacter baumannii, *Pseudomonas aeruginosa*, and *Enterobacteriaceae* (Tacconelli and Magrini, 2017). *Acinetobacter* and *Pseudomonas* were abundant across samples (Figure 1c and Figure S4). Recovering and annotating MAGs for these populations out of AGS biomass could help identify whether they harbour pathogenic traits or not.

Interestingly, the relative abundances of exDNA sequences affiliating with the genera “*Candidatus Microthrix*” (exDNA at beginning of aeration $0.5 \pm 0.1\%$ vs. at end of aeration $5.6 \pm 2.0\%$), *Acidovorax* ($1.2 \pm 0.1\%$ vs. $4.3 \pm 0.2\%$), *Rhodoferrax* ($1.8 \pm 0.2\%$ vs. $7.9 \pm 1.3\%$), and *Zoogloea* ($0.8 \pm 0.2\%$ vs. $2.6 \pm 0.5\%$) increased over the aeration period. These filamentous bacteria (“*Ca. Microthrix*”) and denitrifiers were abundant >3% in the granule fractions of the SBR but not in flocs. These microbial populations are known to populate AGS used for full biological nutrient removal (Winkler et al., 2018).

The presence of exDNA in the water phase can relate to cell lysis or to “active” release from the cells. *Acinetobacter* (iDNA $6.1 \pm 2.8\%$; exDNA $3.4 \pm 2.2\%$), *Flavobacterium* ($1.2 \pm 0.4\%$; $1.0 \pm 0.5\%$), and *Pseudomonas* ($4.3 \pm 0.9\%$; $4.0 \pm 1.1\%$) secrete DNA in the extracellular environment during growth in liquid media (Pietramellara et al., 2009).

Further mechanistic insights, using well-controlled experiments with systems microbiology methods, are required to identify whether cell decay or active secretion can explain the release of exDNA from microorganisms of the sludge.

Figure 1. (a) Schematic representation of an AGS SBR, highlighting biomass metrics and sampling points 1-4. **(b)** PCoA of the absolute microbiome composition at genus-level after normalization and variance stabilization transformation, with Bray-Curtis as distance metric. Colors indicate microbial fractions: iDNA small granule (blue), iDNA large granule (red), iDNA floc (orange), iDNA wastewater (black, including influent and effluent samples), and exDNA (green). Symbols indicate stages in WWT operation: Influent (filled circles), beginning of aeration (asterisks), end of aeration (crossed squares), and effluent (empty circles). **(c)** Taxonomic classification at genus-level heatmap. Colors represent the Z-score computed from the relative abundance at genus-level present above 3% in all classified sequences.

3.3. A wide range of ARGs are present in the effluent exDNA

The resistome of wastewater systems exhibits a large diversity of ARGs. As many as 2,840 ARGs were identified from all the samples, belonging to 15 antibiotic subgroups. ARGs affiliating with MLS

(Macrolide, Lincosamide, and Streptogramin; n = 910 reads), aminoglycosides (n = 598 reads), sulphonamides (n = 375 reads), beta-lactam (n = 330 reads), and tetracycline (n = 233 reads) were the most abundant (**Figure 2a**). The resistome profiles matched with Hendriksen et al. (2019): macrolides are the most abundant ARGs in urban sewage in Europe, followed by aminoglycosides and beta-lactams.

All samples, including the different types of bioaggregates, and iDNA and exDNA fractions, contained ARGs. The exDNA resistome profiles at beginning and end of aeration resembled the iDNA resistomes of granules. Similar to observations made on taxonomic affiliations of exDNA sequences, exARGs mainly originated from the granules rather than from flocs. In the effluent, exDNA was a combination of fragments coming from the influent iDNA (**Figure 2b**) (being in the day 2 part of the bigger dendrogram) plus the different bioaggregates.

The 5 most abundant ARGs over all samples were: aminoglycoside *arr-3 gene* (n = 58 reads), tetracycline *tet(C)* (n = 46 reads), sulfonamide *sul1* (n = 44 reads), multidrug efflux pump *oqxB* (n = 34 reads), and aminoglycoside *aph(3'')-Ib* (n = 34 reads). Taken together (n = 216 reads), these genes only amount to 7.6% of the total ARGs identified. This high diversity of resistance genes matches with another metagenomic analysis of activated sludge samples (Zhou et al., 2019).

Of all 478 ARGs identified in the effluent, 89% were carried by exDNA. This can present a risk for AMR dissemination in the environment and for health by natural transformation of exDNA fragments in microorganisms present in receiving water bodies and drinking water resources, especially for the resistance determinants to the last-resource antibiotics such as carbapenem and colistin.

The types of exARG and iARG were similar, matching with Zhou et al. (2019). In contrast, using PCR only, Li et al. (2020) detected dissimilar compositions of iARGs and exARGs during sludge granulation at pilot scale. Early-stage granulation is a dynamic process involving changing community compositions along the establishment of physicochemical gradients (e.g., increasing anaerobic zones in granules) (Sengar et al., 2018). This can lead to bacteria decay and DNA release shifts over the process, therefore leading to variations in ARG profiles as well. Here, mature granules from the full-scale AGS WWTP were used to track the fate of ARGs under pseudo steady-state conditions.

3.4. Gram-negative bacteria are potential carriers of ARGs in the AGS

All contigs >500 bp containing ARGs were taxonomically classified to identify which microorganisms carried and potentially released ARGs (**Figure 2b; Figure S4**). As high as 98% of contigs containing ARGs were annotated with taxonomy.

A contig affiliated to a taxon does not necessarily mean that the read comes from that specific organism: different populations can share similar ARG sequences. The contigs containing ARGs over the whole metagenomic dataset of all samples and fractions were affiliated to Gram-negative genera like *Pseudomonas* ($8.4 \pm 0.7\%$), *Rhodoferrax* ($5.6 \pm 0.9\%$), *Acinetobacter* ($4.3 \pm 1.4\%$), *Aeromonas* ($3.1 \pm 1.1\%$), *Xanthomonas* ($3.9 \pm 0.9\%$), and *Acidovorax* ($1.9 \pm 0.4\%$) as well as the Gram-positive and antibiotic-producing genus *Streptomyces* ($1.6 \pm 0.4\%$) (**Figure S4**).

Pseudomonas and *Rhodoferrax* genera's high abundance pattern were similar between the two sampling days (**Figure 2b**). iDNA contigs containing ARGs in small and large granules were affiliated at several instances to *Rhodoferrax* ($10.1 \pm 0.4\%$ in granules). Out of the contigs where ARGs were localized, the ARG-containing exDNA affiliated to *Rhodoferrax* (beginning of aeration: $2.0 \pm 0.5\%$; end of aeration: $11.0 \pm 2.9\%$). This increase suggests that DNA was released from *Rhodoferrax* cells during aeration. *Rhodoferrax* is abundant in WWTP effluents and often identified as AMR carriers by encoding multiple efflux pumps (Jin et al., 2020; Zhou et al., 2020). ARGs have been identified from metagenome-assembled genomes (MAGs) retrieved from activated sludge from Danish WWTPs (Miłobedzka et al., 2022; Singleton et al., 2021): *Rhodoferrax* was the most abundant population containing multiple ARGs.

The effluent iDNA is mainly be composed of microorganisms that go through the process without being removed, from bacteria growing on suspended solids and of microorganisms detached from the AGS bioaggregates. Some microorganisms that entered the AGS plant in high abundance and passed through the WWTP are *Aeromonas* and *Acinetobacter*, especially in day 1. The effluent was enriched by microorganisms coming mainly from floccular sludge such as *Lysobacter* (both days) and *Thermomonas* (day 2) (**Figure 2b**). Pathogenic *Arcobacter* crossed WWTPs from influent to effluent (**Figure 1c**) but interestingly, no ARG was detected in any of *Arcobacter* contigs. Further research needs to be done to clearly identify which ARGs in which MGEs are inside specific microbial hosts. Recent analytical advances now allow to identify potential hosts of resistances using Hi-C sequencing (Stalder et al., 2019): populations of *Aeromonadaceae*, *Moraxellaceae*, and *Bacteroidetes* were shown to be critical reservoirs of AMR in WWTPs.

Collectively, the resistome results highlight that AMR determinants released in exDNA during the WWT operation are importantly linked to Gram-negative bacteria, notably *Rhodoferrax*, and to pathogenic bacterial genera carrying ARGs in the effluent such as *Pseudomonas* and *Aeromonas*.

Figure 2. (a) Absolute numbers of classified reads of AR-groups. Labels: (F) floccular aggregates, (LG) large granules, (SG) small granules, (-1) first biological replicates, and (-2) second biological replicates. **(b)** Taxonomic classification at genus-

level of contigs encoding ARGs heatmap. Colors represent the Z-score computed from the relative abundance at genus-level present above 5% in all classified sequences. Labels: (INF) Influent, (BA) Beginning of the aeration, (EA) End of the aeration, and (EFF) effluent.

3.5. MGEs and ARGs often co-localize on contigs recovered from the exDNA of the effluent

Both iDNA and exDNA fractions and all bioaggregates were sources of a diversity of MGEs. The different samples yielded 55,344 different MGEs belonging to bacterial insertion sequences (9,845 reads), integrons, integrases and gene cassettes (20,149 reads), and bacterial integrative conjugative elements (25,350 reads).

When ARGs and MGEs co-localize on the same genetic fragment, there is an increased chance that the fragment can be transferred between bacterial cells. Since MGEs facilitate the transfer, integration, and transposition of genes in genomes, this poses a clear risk for ARG dissemination. From all contigs >500 bp, 312 contigs were identified to contain both ARGs and MGEs (**Table S6**): their 5 most abundant ARGs were *blaVIM-48* (28), *cmx-1* (28), *sul2* (25), *tetA* (24), and *aadA6* (24). A set of 80 contigs contained multiple ARGs, and of these, 60 also contained MGEs: their 5 most abundant ARGs were all aminoglycoside resistance genes, namely *aph(3'')-Ib* (17), *aadA6* (17), *aph(6)-Id* (17), *aadA11* (7), and *aac(3)-Ib* (7).

Multiple co-localization patterns were detected only in the iDNA of granules and in the exDNA of the effluent. For instance, *sul2* and *tetA* were exclusively found in the iDNA from large granules (replicate 2) from the end of the aeration, and in the effluent exDNA fraction (replicate 2). This contig also aligned with ISVs3, an insertion sequence often found on plasmids (Toleman and Walsh, 2010).

Of the 60 contigs >500bp were multiple ARGs and MGEs co-localized, 3 major groups of contigs are highlighted hereafter because of their abundance in the system or the risk associated with leaving the WWTP in the exDNA of the effluent.

Firstly, as shown in **Table S6**, 15 contigs were identified both as carriers of two aminoglycoside resistance genes (*aph(3'')-Ib* and *aph(6)-Id*), and as part of transposon Tn5393 (**Figure 3a**). These contigs were found in all samples across WWTP operation for all microbial aggregates and exDNA and iDNA fractions (excluding the flocs). After alignment to the NCBI nucleotide database, the contigs aligned with IncQ plasmids, mostly from populations of *Acinetobacter* and *Pseudomonas* genera, previously identified in our wastewater samples. IncQ plasmids are significant since they are related

to the broadest host range of all replicating elements in bacteria (Meyer, 2009). They are known to mobilize via conjugation although natural transformation has also been observed (Bönemann et al., 2006; Meyer, 2009; Szczepanowski et al., 2009, 2004).

Then, 18 contigs aligned with a Tn3 family transposon (TnAs3), normally encoding a β -lactamase gene, a transposase and a resolvase (Nicolas et al., 2015) as identified from the exDNA effluent fraction (**Figure 3b**). These contigs (identities >95%) aligned exclusively with plasmids from *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Enterobacter cloacae*, *Enterobacter kobei*, and *Klebsiella oxytoca*. Interestingly, of all pathogens, the World Health Organization stressed that addressing carbapenem-resistance in *Pseudomonas aeruginosa* and *Enterobacteriaceae* should have the highest priority (Tacconelli and Magrini, 2017).

Finally, another contig of importance found in the exDNA fraction of the effluent carried multiple resistance genes against aminoglycosides (*aac(6')*), beta-lactamases (*blaOXA-10*), and phenicol (*cmlA1*) (**Figure 3c**). This contig aligned with a 100% sequence identity to a plasmid already annotated from *Aeromonas hydrophila* (pWCX23_1), carrying 15 different ARGs (Nwaiwu and Aduba, 2020).

Figure 3. Contigs containing both ARGs and MGEs. **(a)** Contig of length 5595 bp, containing the class 1 integron-integrase (*int1*), aminoglycoside phosphotransferase (*aph(6)-Id*), also known as *strB*, an aminoglycoside phosphotransferase (*aph(3'')*-*Ib* also known as *strA*), a sulfonamide resistant dihydropteroate synthase (*sul1*), and a fully classified transposon Tn5393. **(b)** Contig of length 1754 bp, containing the class 1 integron-integrase (*int1*), beta-lactamase *bla-GES-5*, integron-encoded aminoglycoside acetyltransferase fusion protein *ant(3'')-li-aac(6')-IId*, fully classifying as transposon TnAs3. **(c)** Contig of length 2594 bp, containing beta-lactamase *blaOXA-10*, a plasmid or transposon-encoded chloramphenicol exporter gene *cmlA1*, an integron-encoded aminoglycoside acetyltransferase *aac(6')-Ib*, and a transposase ISAb43.

The strong presence of mobilized resistance determinants in the extracellular fraction of the AGS WWTP has been also described by Ikuma and Rehmann (2020). Using mathematical modelling, they suggested that, at the WWTP effluent discharge point in a river, the total number count of ARGs was 13 times higher when including exARGs on top of iARGs. Collectively, exARGs in WWTP discharges need to be taken into account for accurate risk assessments of AMR.

Co-localization studies are valuable tools to identify the genetic structures and potential hosts by which resistance determinants are carried and transferred between bacteria. More research needs to be conducted to reconstruct full plasmids from wastewater effluents. This will shed light on the type of plasmids released into the environment, *i.e.*, narrow/broad-range and synthetic/naturally occurring, and on the genetic constructions ARGs are embedded in.

3.6.No antibiotic-related biosynthetic gene clusters were found in the biomass for the most abundant ARG identified

Functional and pathway analyses were performed on the contigs to identify AMR pathways and on the MAGs to identify potential intrinsic antibiotic production by microorganisms that could promote the development and survival of ARB in AGS systems.

From all the AMR pathways available in the KEGG database (Kanehisa et al., 2002), we identified the cationic antimicrobial peptide (CAMP) resistance pathway, vancomycin resistance pathway, and beta-lactamase resistance pathway, including efflux pumps that actively transport the beta-lactam antibiotics out of the cell (RND efflux pumps) (**Figure S6-S8**). Both influent and effluent iDNA and exDNA were sources of genes involved in beta-lactam resistance. Most genes involved in these resistance pathways were present in the influent iDNA but not in the influent exDNA. However, in the effluent, both iDNA and exDNA contained resistance pathways genes.

From the antiSMASH analysis of MAGs recovered from small and large granules (no MAGs could be recovered from flocs due to the sequencing depth and short reads used) at the end of aeration, positive hits on non-ribosomal peptide synthetase (NRPS) biosynthetic gene clusters (BGCs) were detected in MAGs of *Nitrospira* and *Janthinobacterium*. NRPS antibiotics include vancomycin, polymyxin, and teixobactin, which are highly effective against multidrug-resistant bacteria (Li et al., 2018). The vancomycin resistance pathway was one of the few reconstructed pathways in the samples analyzed (**Figure S6**). Terpenes, indoles, T1PKS, and ectoine BGCs were identified as the most recurrent (**Table S8**). No aminoglycosides or beta-lactams antibiotic BGCs were found in any of the DNA samples.

3.7. ARGs increase in exDNA over the WWT operation

To quantify the ARG and ARB occurrence and related removal capacity of the AGS WWTP, a panel of six ARGs and *int11* as MGE were quantified by qPCR. Two normalization methods (using *16S rRNA* and *rpoB* genes) were assessed. Both the ARGs and the MGE were consistently detected across all the AGS process phases and bioaggregates.

The most predominant genes throughout the process were the sulfonamide resistance gene *sul1* and the class 1 integron gene *int11* (**Figure 4a**). These genes were the most concentrated ones in the effluent with 7.5 (*sul1*) and 7.0 (*int11*) $\log_{10}$ gene copies mL^{-1} (sum of iDNA and exDNA pools). Genes with the lowest concentration were *bla_{CTXM}* (2.1), *qnrS* (3.1), and *tetO* (3.2) (same units as above). Absolute and normalized concentrations of ARGs and MGEs, and iDNA vs. exDNA analysis, are shown in **Figure S9** and **Table S7**. These ARGs concentrations are in agreement with recent studies (Calderón-Franco et al., 2020; Pallares-Vega et al., 2019) and comparable to others, especially when considering the variance in sewage and treatment efficiency (Alexander et al., 2020; Lee et al., 2017; Wang et al., 2020).

The extended-spectrum beta-lactamase (ESBL) gene (*bla_{CTXM}*) was detected only in the influent ($3.0 \pm 0.6 \log_{10}$ gene copies mL^{-1}) and effluent ($2.1 \pm 0.1 \log_{10}$ gene copies mL^{-1}) iDNAs. No copies of this gene were found during the aeration process or in the exDNA fractions. In the most recent NethMap/MARAN report presenting data on antibiotics use and resistance in the Netherlands (RIVM and SWAB, 2020), the increasing detection of bacteria producing ESBL, even in countries with non-extensive use of antibiotics, was highlighted as an imminent threat. The low but increasing presence of ESBL-producing bacteria can justify the small but significant detection of beta-lactam variants such as *bla_{CTXM}*.

3.8. AGS treatment is efficient at reducing the load of resistance determinants

The absolute concentration of iARGs was significantly higher ($p < 0.05$ - 0.005) than of exARGs across all SBR phases and bioaggregates (**Figure 4a**). The average iDNA/exDNA ratio was 1.90 in the influent and 2.3 in the effluent (**Figure S9**). These differences in the concentration of iARGs and exARGs match with a recent report of highest relative abundance of sulfonamide, tetracycline resistance, and integron genes in the iDNA than in the exDNA of an AGS WWTP (Li et al., 2020). We found a significant reduction ($p < 0.005$) after treatment for all genes analyzed by qPCR in the iDNA fraction. In contrast, multiple genes such as the bacterial proxy genes *16S rRNA* and *rpoB* as well as the resistance genes *ermB* ($2.1 \log_{10}$ gene copies mL^{-1} higher) and *sul1* ($1.0 \log_{10}$ gene copies mL^{-1} higher) increase from the

influent to effluent exDNA (**Figure S11**). Di Cesare et al. (2016) have reported different ARGs patterns of exARGs, iARGs, and integrase MGEs: microbial cell decay and cell lysis can potentially release genetic material extracellularly. The free-floating exDNA is subjected to the fluctuating conditions of the WWTP environment differently than iDNA (Li et al., 2020). Its ability to maintain mobility and be involved in resistance spread needs to be studied in the future.

Figure 4. (a) Concentration of selected ARGs and *int1* across AGS WWT phases. Error bars represent the standard deviation between biological replicates (n=3). The significances of the difference between exDNA and iDNA measurements are indicated by the Wilcoxon test and displayed as p<0.05 (*) and p<0.005 (**). **(b)** Absolute abundance of ARGs and *int1* before and after treatment (influent and effluent phase) expressed in log₁₀ gene copies mL⁻¹. **(c)** Relative abundance of genes in the influent and effluent expressed as log₁₀ gene copies / log₁₀ 16S rRNA copies. The significance of the difference between influent and effluent measurements, as indicated by the Wilcoxon test, is displayed as p<0.05 (*).

Small granules accumulated higher concentrations of resistance genes (average 0.82 ± 0.31 $\log_{10}$ gene copies/ $\log_{10}$ *16S rRNA* copies) in their iDNA than big granules (0.62 ± 0.26) or flocs (0.61 ± 0.25) at the end of the aeration. No significant differences were observed between granules at the beginning of the aeration (**Figure 4a; Figure S9**). It is not clear why small granules accumulated more ARGs during the 9-h aeration process.

Flocs and small granules are more susceptible to immigration from the influent/sewage than large granules (Ali et al., 2019). The solid retention time (SRT) of small granules is much shorter (7.7 ± 0.5 days) than that of big granules (142.6 ± 14.9 days) (Ali et al., 2019). If the influent is enriched with ARB, immigration can explain a faster loading of ARGs in small than large granules. By being more rapidly released out of the tank, small granules can play a role in the release and dissemination of ARGs through WWTP effluent than big granules. Because of their higher specific surface area, fractions of small granules can capture more free-floating exDNA released from cell decay during the treatment process. exDNA can help stabilize their architecture by adhesion at initial granulation stages through acid-base interactions (Wang et al., 2020; Das et al., 2011, 2010).

Further studies on ARG distribution across bioaggregates are needed to determine if ARGs are homogeneously distributed across the cross-section of the granules or mainly located at their surface or in one of their layers. Fluorescent *in situ* hybridization with ARG probes could determine the ARG distribution in granule slices. Such studies will also indicate if transfer between ARGs is more prone to happen between or within granules.

3.9. Resistance loads in wastewater effluents depend on DNA fraction

Looking at the removal capacity of the full process from influent to effluent, the overall pool of resistance determinants was significantly reduced during WWT (average $1.1 \log_{10}$ gene copies mL^{-1} removal). Some ARGs were still discharged with iDNA inside cells ($2\text{-}4 \log_{10}$ gene copies g VSS^{-1}) and with exDNA free-floating in water ($0.5\text{-}3 \log_{10}$ gene copies mL^{-1}) (**Figure 4a; Figure S9**). The majority of the targeted genes decreased after the treatment. From the panel analyzed by qPCR, *int11*, *bla_{CTXM}*, and *tetO* significantly ($p < 0.05$) decreased by 1.3, 1.7, and $2.3 \log_{10}$ gene copies mL^{-1} , respectively. The AGS process is partially effective at reducing ARGs: Sabri et al. (2020) have reported a similar drop for tetracycline resistance gene (*tetW*) in an AGS technology of around $2 \log_{10}$ gene copies mL^{-1} .

Based on the remaining ARGs concentrations in the effluent, an average of $6.3 \log_{10}$ gene copies mL^{-1} were released in total, iDNA plus exDNA (**Figure 4b**). Taking hospital discharges as references, this can be considered high ($>10^4$ gene copies mL^{-1}) (Le et al., 2016). From **Figure 4c**, the normalized values did not significantly enrich bacterial populations carrying ARGs or integrases.

Overall, we found that the AGS WWTP can reduce the amount of ARB but still releases in the environment a significant amount of ARGs enclosed on MGEs of free-floating exDNA. The commonly overlooked exDNA should therefore be considered (together with the iDNA of ARB) as an important factor in the dissemination of resistance determinants in the aquatic environment. Risks associated with exDNA and iDNA fractions discharged in water resources need to be evaluated, with a clear identification of exposures and effects.

4. Conclusions

This work led to the following main conclusions:

1. *Pseudomonas* and *Rhodospirillum* populations were the main hosts of antibiotic resistance genes (ARGs) in the biomass. Other ARG-carrying populations affiliated with *Acinetobacter*, *Aeromonas*, *Xanthomonas*, *Acidovorax*, *Bacteroidetes*, and *Streptomyces*.
2. Several ARGs co-localized with mobile genetic elements (MGEs) on genetic contigs of free-floating extracellular DNA, thus being potentially transferrable to microorganisms. The most abundant ARGs co-localizing with MGEs were the beta-lactamases resistance gene *blaVIM-48*, chloramphenicol resistance gene *cmx-1*, sulfonamides resistance gene *sul2*, tetracycline resistance gene *tetA*, and aminoglycoside resistance gene *aadA6*.
3. The panel of ARGs and *int1* MGE analyzed by qPCR were detected in all samples.
4. The ARG fraction in intracellular DNA decreased with $1.1 \log_{10}$ gene copies mL^{-1} during the process. Conversely, some ARGs located in the exDNA (*ermB*, *sul1* and *sul2*) increased during the process.
5. According to metagenomics, exDNA carries ARGs enclosed inside a diversity of MGEs, detected in the WWTP effluent. These can potentially spread AMR in aquatic environments by horizontal gene transfer.

Therefore, studies for the surveillance, risk assessment, and mitigation of AMR in wastewater environments should consider not only iDNA but also exDNA pools.

Data availability

Metagenome sequencing data were deposited in the NCBI database with the BioProject ID PRJNA783874.

Conflict of interest statement

The authors declare no conflict of interest.

Authors' contributions

DCF designed the study with inputs of MvL, MP, TA, and DGW. DCF, RS, and SC performed the experimental investigations. DCF wrote the manuscript with direct contribution, edits, and critical feedback by all authors.

Acknowledgements

We are grateful to Pascalle Vermeulen from Royal HaskoningDHV, The Netherlands, for helping us arrange and design this sampling campaign. This work is part of the research project "Transmission of Antimicrobial Resistance Genes and Engineered DNA from Transgenic Biosystems in Nature" (TARGETBIO) funded by the programme Biotechnology & Safety Program of the Ministry of Infrastructure and Water Management (grant no. 15812) of the Applied and Engineering Sciences (TTW) Division of the Dutch Research Council (NWO).

References

- Abe, K., Nomura, N., Suzuki, S., 2020. Biofilms: Hot spots of horizontal gene transfer (HGT) in aquatic environments, with a focus on a new HGT mechanism. *FEMS Microbiol. Ecol.* <https://doi.org/10.1093/FEMSEC/FIAA031>
- Alexander, J., Hembach, N., Schwartz, T., 2020. Evaluation of antibiotic resistance dissemination by wastewater treatment plant effluents with different catchment areas in Germany. *Sci. Rep.* 10, 1–9. <https://doi.org/10.1038/s41598-020-65635-4>
- Ali, M., Wang, Z., Salam, K.W., Hari, A.R., Pronk, M., Van Loosdrecht, M.C.M., Saikaly, P.E., 2019. Importance of species sorting and immigration on the bacterial assembly of different-sized aggregates in a full-scale Aerobic granular sludge plant. *Environ. Sci. Technol.* 53, 8291–8301. <https://doi.org/10.1021/acs.est.8b07303>
- Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.
- Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T., 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, 29–35. <https://doi.org/10.1093/nar/gkab335>
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bönemann, G., Stiens, M., Pühler, A., Schlüter, A., 2006. Mobilizable IncQ-related plasmid carrying a new quinolone resistance gene, qnrS2, isolated from the bacterial community of a wastewater treatment plant. *Antimicrob. Agents Chemother.* 50, 3075–3080. <https://doi.org/10.1128/AAC.00378-06>
- Bortolaia, V., Kaas, R.S., Ruppe, E., Roberts, M.C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R.L., Rebelo, A.R., Florensa, A.F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J.K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B.B., Malhotra-Kumar, S., Westh, H., Pinholt, M., Anjum, M.F., Duggett, N.A., Kempf, I., Nykäsenoja, S., Olkkola, S., Wiczorek, K., Amaro, A., Clemente, L., Mossong, J., Losch, S., Ragimbeau, C., Lund, O., Aarestrup, F.M., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* 75. <https://doi.org/10.1093/jac/dkaa345>
- Breitwieser, F.P., Salzberg, S.L., 2020. Pavian: Interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* 36, 1303–1304. <https://doi.org/10.1093/bioinformatics/btz715>
- Bürgmann, H., Frigon, D., H Gaze, W., M Manaia, C., Pruden, A., Singer, A.C., F Smets, B., Zhang, T., 2018. Water and sanitation: an essential battlefront in the war on antimicrobial resistance. *FEMS Microbiol. Ecol.* 94, 1–14. <https://doi.org/10.1093/femsec/fiy101>
- Calderón-Franco, D., van Loosdrecht, M., Abeel, T., Weissbrodt, D., 2020. Free-floating extracellular DNA: Systematic profiling of mobile genetic elements and antibiotic resistance from wastewater. *Water Res.* 189, 1–13. <https://doi.org/10.1101/2020.05.01.072397>
- Das, T., Sharma, P.K., Busscher, H.J., Van Der Mei, H.C., Krom, B.P., 2010. Role of extracellular DNA in initial bacterial adhesion and surface aggregation. *Appl. Environ. Microbiol.* 76, 3405–3408. <https://doi.org/10.1128/AEM.03119-09>
- Das, T., Sharma, P.K., Krom, B.P., Van Der Mei, H.C., Busscher, H.J., 2011. Role of eDNA on the adhesion forces between streptococcus mutans and substratum surfaces: Influence of ionic strength and substratum hydrophobicity. *Langmuir* 27, 10113–10118. <https://doi.org/10.1021/la202013m>
- Davies, J., Davies, D., 2010. Microbiology and Molecular Biology Reviews Microbiology and Molecular Biology Reviews Origins and Evolution of Antibiotic Resistance. *Mol. Biol. Rev. Microbiol. Mol. Biol. Rev.* 74.
- Di Cesare, A., Eckert, E.M., D’Urso, S., Bertoni, R., Gillan, D.C., Wattiez, R., Corno, G., 2016. Co-occurrence of integrase 1, antibiotic and heavy metal resistance genes in municipal wastewater treatment plants. *Water Res.* 94, 208–214.

<https://doi.org/10.1016/j.watres.2016.02.049>

Dixon, P., 2003. Computer program review VEGAN , a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930.

Dong, P., Wang, H., Fang, T., Wang, Y., Ye, Q., 2019. Assessment of extracellular antibiotic resistance genes (eARGs) in typical environmental samples and the transforming ability of eARG. *Environ. Int.* 125, 90–96. <https://doi.org/10.1016/j.envint.2019.01.050>

Guo, H., van Lier, J.B., de Kreuk, M., 2020. Digestibility of waste aerobic granular sludge from a full-scale municipal wastewater treatment system. *Water Res.* 173, 115617. <https://doi.org/10.1016/j.watres.2020.115617>

Guo, J., Li, J., Chen, H., Bond, P.L., Yuan, Z., 2017. Metagenomic analysis reveals wastewater treatment plants as hotspots of antibiotic resistance genes and mobile genetic elements. *Water Res.* 123, 468–478. <https://doi.org/10.1016/j.watres.2017.07.002>

Hendriksen, R.S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S.K., Kjeldgaard, J., Kaas, R.S., Clausen, P.T.L.C., Vogt, J.K., Leekitcharoenphon, P., van de Schans, M.G.M., Zuidema, T., de Roda Husman, A.M., Rasmussen, S., Petersen, B., Bego, A., Rees, C., Cassar, S., Coventry, K., Collignon, P., Allerberger, F., Rahube, T.O., Oliveira, G., Ivanov, I., Vuthy, Y., Sopheak, T., Yost, C.K., Ke, C., Zheng, H., Baisheng, L., Jiao, X., Donado-Godoy, P., Coulibaly, K.J., Jergović, M., Hrenovic, J., Karpíšková, R., Villacis, J.E., Legesse, M., Eguale, T., Heikinheimo, A., Malania, L., Nitsche, A., Brinkmann, A., Saba, C.K.S., Kocsis, B., Solymosi, N., Thorsteinsdottir, T.R., Hatha, A.M., Alebouyeh, M., Morris, D., Cormican, M., O'Connor, L., Moran-Gilad, J., Alba, P., Battisti, A., Shakenova, Z., Kiiyukia, C., Ng'eno, E., Raka, L., Avsejenko, J., Bērziņš, A., Bartkevics, V., Penny, C., Rajandas, H., Parimannan, S., Haber, M.V., Pal, P., Jeunen, G.J., Gemmell, N., Fashae, K., Holmstad, R., Hasan, R., Shakoar, S., Rojas, M.L.Z., Wasyl, D., Bosevska, G., Kochubovski, M., Radu, C., Gassama, A., Radosavljevic, V., Wuertz, S., Zuniga-Montanez, R., Tay, M.Y.F., Gavačová, D., Pastuchova, K., Truska, P., Trkov, M., Esterhuyse, K., Keddy, K., Cerdà-Cuellar, M., Pathirage, S., Norrgren, L., Örn, S., Larsson, D.G.J., Heijden, T., Van der Kumburu, H.H., Sanneh, B., Bidjaja, P., Njanpop-Lafourcade, B.M., Nikiema-Pessinaba, S.C., Levent, B., Meschke, J.S., Beck, N.K., Van, C.D., Phuc, N. Do, Tran, D.M.N., Kwenda, G., Tabo, D. adjim, Wester, A.L., Cuadros-Orellana, S., Amid, C., Cochrane, G., Sicheritz-Ponten, T., Schmitt, H., Alvarez, J.R.M., Aidara-Kane, A., Pamp, S.J., Lund, O., Hald, T., Woolhouse, M., Koopmans, M.P., Vigre, H., Petersen, T.N., Aarestrup, F.M., 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* 10. <https://doi.org/10.1038/s41467-019-08853-3>

Ikuma, K., Rehmann, C.R., 2020. Importance of extracellular DNA in the fate and transport of antibiotic resistance genes downstream of a wastewater treatment plant. *Environ. Eng. Sci.* 37, 164–168. <https://doi.org/10.1089/ees.2019.0319>

Jin, C.Z., Zhuo, Y., Wu, X., Ko, S.R., Li, T., Jin, F.J., Ahn, C.Y., Oh, H.M., Lee, H.G., Jin, L., 2020. Genomic and metabolic insights into denitrification, sulfur oxidation, and multidrug efflux pump mechanisms in the bacterium *rhodoferax sediminis* sp. Nov. *Microorganisms* 8, 5–7. <https://doi.org/10.3390/microorganisms8020262>

Kanehisa, M., Goto, S., Kawashima, S., Nakaya, A., 2002. The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46. <https://doi.org/10.1093/nar/30.1.42>

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z., 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019, 1–13. <https://doi.org/10.7717/peerj.7359>

Kolde, R., 2019. Package ‘pheatmap’: pretty heat map 1–8.

Kristensen, J.M., Nierychlo, M., Albertsen, M., Nielsen, P.H., 2020. Bacteria from the genus *arcobacter* are abundant in effluent from wastewater treatment plants. *Appl. Environ. Microbiol.* 86. <https://doi.org/10.1128/AEM.03044-19>

- Le, T.H., Ng, C., Chen, H., Yi, X.Z., Koh, T.H., Barkham, T.M.S., Zhou, Z., Gin, K.Y.H., 2016. Occurrences and characterization of antibiotic-resistant bacteria and genetic determinants of hospital wastewater in a tropical country. *Antimicrob. Agents Chemother.* 60, 7449–7456. <https://doi.org/10.1128/AAC.01556-16>
- Lee, J., Jeon, J.H., Shin, J., Jang, H.M., Kim, S., Song, M.S., Kim, Y.M., 2017. Quantitative and qualitative changes in antibiotic resistance genes after passing through treatment processes in municipal wastewater treatment plants. *Sci. Total Environ.* 605–606, 906–914. <https://doi.org/10.1016/j.scitotenv.2017.06.250>
- Li, D., Gao, J., Dai, H., Duan, W., Wang, Z., Zhou, Z., 2020. Fates of intracellular and extracellular antibiotic resistance genes during a pilot-scale aerobic granular sludge cultivation process. *Chem. Eng. J.* <https://doi.org/10.1016/j.cej.2020.127737>
- Li, Y., Zhong, Z., Hou, P., Zhang, W.-P., Qian, P.-Y., 2018. Resistance to nonribosomal peptide antibiotics mediated by D-stereospecific peptidases. *Nat. Chem. Biol.* 14, 381–387. <https://doi.org/10.1038/s41589-018-0009-4>
- Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., Tai, C., Deng, Z., Ou, H.Y., 2019. ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. <https://doi.org/10.1093/nar/gky1123>
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. <https://doi.org/10.1186/s13059-014-0550-8>
- Ma, L., Li, A.D., Yin, X. Le, Zhang, T., 2017. The Prevalence of Integrins as the Carrier of Antibiotic Resistance Genes in Natural and Man-Made Environments. *Environ. Sci. Technol.* 51, 5721–5728. <https://doi.org/10.1021/acs.est.6b05887>
- Madsen, J.S., Burmølle, M., Hansen, L.H., Sørensen, S.J., 2012. The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunol. Med. Microbiol.* 65, 183–195. <https://doi.org/10.1111/j.1574-695X.2012.00960.x>
- Meyer, F., Bagchi, S., Chaterji, S., Gerlach, W., Grama, A., Harrison, T., Paczian, T., Trimble, W.L., Wilke, A., 2018. MG-RAST version 4-Lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. *Brief. Bioinform.* 20, 1151–1159. <https://doi.org/10.1093/bib/bbx105>
- Meyer, R., 2009. Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* 62, 57–70. <https://doi.org/10.1016/j.plasmid.2009.05.001>
- Mitobedzka, A., Ferreira, C., Vaz-Moreira, I., Calderón-Franco, D., Gorecki, A., Purkrtova, S., Jan Bartacek, Dziewit, L., Singleton, C.M., Nielsen, P.H., Weissbrodt, D.G., Manaia, C.M., 2022. Monitoring antibiotic resistance genes in wastewater environments: The challenges of filling a gap in the One-Health cycle. *J. Hazard. Mater.* 424. <https://doi.org/10.1016/j.jhazmat.2021.127407>
- Moore, D., Dowhan, D., 2002. Purification and Concentration of DNA. *Curr. Protoc. Mol. Biol.*
- Moura, A., Soares, M., Pereira, C., Leitão, N., Henriques, I., Correia, A., 2009. INTEGRALL: A database and search engine for integrins, integrases and gene cassettes. *Bioinformatics* 25, 1096–1098. <https://doi.org/10.1093/bioinformatics/btp105>
- Nicolas, E., Lambin, M., Dandoy, D., Galloy, C., Nguyen, N., Oger, C.A., Hallet, B., 2015. The Tn3-family of replicative transposons. *Mob. DNA III* 693–726. <https://doi.org/10.1128/9781555819217.ch32>
- Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Jiang, C., Albertsen, M., Dueholm, M.S., Nielsen, P.H., 2020. MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge. *Water Res.* 182, 115955. <https://doi.org/10.1016/j.watres.2020.115955>
- Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 27, 824–834. <https://doi.org/10.1101/gr.213959.116>

964 Nwaiwu, O., Aduba, C.C., 2020. An in silico analysis of acquired antimicrobial resistance genes in aeromonas plasmids.
965 AIMS Microbiol. 6, 75–91. <https://doi.org/10.3934/microbiol.2020005>

966 Pallares-Vega, R., Blaak, H., van der Plaats, R., de Roda Husman, A.M., Hernandez Leal, L., van Loosdrecht, M.C.M.,
967 Weissbrodt, D.G., Schmitt, H., 2019. Determinants of presence and removal of antibiotic resistance genes during
968 WWTP treatment: A cross-sectional study. Water Res. 161, 319–328. <https://doi.org/10.1016/j.watres.2019.05.100>

969 Pallares-Vega, R., Hernandez Leal, L., Fletcher, B.N., Vias-Torres, E., van Loosdrecht, M.C.M., Weissbrodt, D.G., Schmitt, H.,
970 2021. Annual dynamics of antimicrobials and resistance determinants in flocculent and aerobic granular sludge
971 treatment systems. Water Res. 190, 116752. <https://doi.org/10.1016/j.watres.2020.116752>

972 Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M.T., Guerri, G., Nannipieri, P., 2009. Extracellular DNA in soil and
973 sediment: Fate and ecological relevance. Biol. Fertil. Soils 45, 219–235. <https://doi.org/10.1007/s00374-008-0345-8>

974 Pronk, M., de Kreuk, M.K., de Bruin, B., Kamminga, P., Kleerebezem, R., van Loosdrecht, M.C.M., 2015. Full scale
975 performance of the aerobic granular sludge process for sewage treatment. Water Res. 84, 207–217.
976 <https://doi.org/10.1016/j.watres.2015.07.011>

977 R Foundation for Statistical Computing., 2018. R: a Language and Environment for Statistical Computing., [http://www.R-](http://www.R-project.org/)
978 [project.org/](http://www.R-project.org/).

979 RIVM, SWAB, 2020. NethMap 2020.

980 Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of
981 digital gene expression data. Bioinformatics 26, 139–140. <https://doi.org/10.1093/bioinformatics/btp616>

982 Sabri, N.A., van Holst, S., Schmitt, H., van der Zaan, B.M., Gerritsen, H.W., Rijnaarts, H.H.M., Langenhoff, A.A.M., 2020. Fate
983 of antibiotics and antibiotic resistance genes during conventional and additional treatment technologies in
984 wastewater treatment plants. Sci. Total Environ. 741, 140199. <https://doi.org/10.1016/j.scitotenv.2020.140199>

985 Seemann, T., 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069.
986 <https://doi.org/10.1093/bioinformatics/btu153>

987 Sengar, A., Basheer, F., Aziz, A., Farooqi, I.H., 2018. Aerobic granulation technology: Laboratory studies to full scale
988 practices. J. Clean. Prod. 197, 616–632. <https://doi.org/10.1016/j.jclepro.2018.06.167>

989 Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., Chandler*, and M., 2006. ISfinder: the reference centre for bacterial
990 insertion sequences. Nucleic Acids Res. 34, D32–D36. <https://doi.org/10.1093/nar/gkj014>

991 Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaitė, Z., Karst,
992 S.M., Dueholm, M.S., Nielsen, P.H., Albertsen, M., 2021. Connecting structure to function with the recovery of over
993 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nat.
994 Commun. 12, 1–13. <https://doi.org/10.1038/s41467-021-22203-2>

995 Slizovskiy, I.B., Mukherjee, K., Dean, C.J., Boucher, C., Noyes, N.R., 2020. Mobilization of Antibiotic Resistance: Are Current
996 Approaches for Colocalizing Resistomes and Mobilomes Useful? Front. Microbiol. 11, 1–20.
997 <https://doi.org/10.3389/fmicb.2020.01376>

998 Stalder, T., Press, M.O., Sullivan, S., Liachko, I., Top, E.M., 2019. Linking the resistome and plasmidome to the microbiome.
999 ISME J. 2437–2446. <https://doi.org/10.1038/s41396-019-0446-4>

1000 Szczepanowski, R., Krahn, I., Linke, B., Goesmann, A., Pühler, A., Schlüter, A., 2004. Antibiotic multiresistance plasmid
1001 pRSB101 isolated from a wastewater treatment plant is related to plasmids residing in phytopathogenic bacteria and
1002 carries eight different resistance determinants including a multidrug transport system. Microbiology 150, 3613–
1003 3630. <https://doi.org/10.1099/mic.0.27317-0>

1004 Szczepanowski, R., Linke, B., Krahn, I., Gartemann, K.H., Gützkow, T., Eichler, W., Pühler, A., Schlüter, A., 2009. Detection of
1005 140 clinically relevant antibiotic-resistance genes in the plasmid metagenome of wastewater treatment plant

bacteria showing reduced susceptibility to selected antibiotics. *Microbiology* 155, 2306–2319.
<https://doi.org/10.1099/mic.0.028233-0>

Tacconelli, E., Magrini, N., 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics., World Health Organization.

Toh, S.K., Tay, J.H., Moy, B.Y.P., Ivanov, V., Tay, S.T.L., 2003. Size-effect on the physical characteristics of the aerobic granule in a SBR. *Appl. Microbiol. Biotechnol.* 60, 687–695. <https://doi.org/10.1007/s00253-002-1145-y>

Toleman, M.A., Walsh, T.R., 2010. ISCR Elements Are Key Players in IncA/C Plasmid Evolution. *Antimicrob. Agents Chemother.* 54, 3534. <https://doi.org/10.1128/AAC.00383-10>

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N., 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903.
<https://doi.org/10.1038/nmeth.3589>

Ukuhor, H.O., 2021. The interrelationships between antimicrobial resistance, COVID-19, past, and future pandemics. *J. Infect. Public Health.* <https://doi.org/10.1016/j.jiph.2020.10.018>

van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.

Wang, J., Chu, L., Wojnárovits, L., Takács, E., 2020. Occurrence and fate of antibiotics, antibiotic resistant genes (ARGs) and antibiotic resistant bacteria (ARB) in municipal wastewater treatment plant: An overview. *Sci. Total Environ.* 744, 140997. <https://doi.org/10.1016/j.scitotenv.2020.140997>

Wang, Y. qiao, Li, W., Zhuang, J. long, Liu, Y. di, Shapleigh, J.P., 2020. Bacteriophage-mediated extracellular DNA release is important for the structural stability of aerobic granular sludge. *Sci. Total Environ.* 726, 138392.
<https://doi.org/10.1016/j.scitotenv.2020.138392>

Weissbrodt, D.G., Shani, N., Holliger, C., 2014. Linking bacterial population dynamics and nutrient removal in the granular sludge biofilm ecosystem engineered for wastewater treatment. *FEMS Microbiol. Ecol.* 88, 579–595.
<https://doi.org/10.1111/1574-6941.12326>

WHO & UN, 2019. No time to wait: Securing the future from drug-resistant infections report to the secretary-general of the United Nations.

Winkler, M.K.H., Meunier, C., Henriot, O., Mahillon, J., Suárez-Ojeda, M.E., Del Moro, G., De Sanctis, M., Di Iaconi, C., Weissbrodt, D.G., 2018. An integrative review of granular sludge for the biological removal of nutrients and recalcitrant organic matter from wastewater. *Chem. Eng. J.* 336, 489–502. <https://doi.org/10.1016/j.cej.2017.12.026>

Wood, D.E., Lu, J., Langmead, B., 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 1–13.
<https://doi.org/10.1101/762302>

Yuan, Q. Bin, Huang, Y.M., Wu, W. Bin, Zuo, P., Hu, N., Zhou, Y.Z., Alvarez, P.J.J., 2019. Redistribution of intracellular and extracellular free & adsorbed antibiotic resistance genes through a wastewater treatment plant by an enhanced extracellular DNA extraction method with magnetic beads. *Environ. Int.* 131, 104986.
<https://doi.org/10.1016/j.envint.2019.104986>

Zhang, T., Zhang, X.-X., Ye, L., 2011. Plasmid Metagenome Reveals High Levels of Antibiotic Resistance Genes and Mobile Genetic Elements in Activated Sludge. *PLoS One* 6, 26041. <https://doi.org/10.1371/journal.pone.0026041>

Zhou, C.S., Wu, J.W., Dong, L.L., Liu, B.F., Xing, D.F., Yang, S.S., Wu, X.K., Wang, Q., Fan, J.N., Feng, L.P., Cao, G.L., 2020. Removal of antibiotic resistant bacteria and antibiotic resistance genes in wastewater effluent by UV-activated persulfate. *J. Hazard. Mater.* 388, 122070. <https://doi.org/10.1016/j.jhazmat.2020.122070>

Zhou, S., Zhu, Y., Yan, Y., Wang, W., Wang, Y., 2019. Deciphering extracellular antibiotic resistance genes (eARGs) in activated sludge by metagenome. *Water Res.* 161, 610–620. <https://doi.org/10.1016/j.watres.2019.06.048>

Supplementary material

Quantitative PCR conditions, primers and standards

Quantitative polymerase chain reaction experiments were performed for the quantification of the genes in **Table S1** in the WWTP's DNA samples (Section 2.1) using the thermal cycler qTOWER3 (Analytica Jena, Germany). Based on the measured cycle threshold values (Ct values), standard curves were derived, which indicated the efficiency of the reaction and enabled further determination of the genes' concentration. The DNA standards sequences, the standard curves, the reaction results per qPCR, and the calculation procedure are mentioned in Appendix A. The primers used for each gene and the amplified fragments' lengths are listed in **Table S1**. The qPCR analysis was performed in a total volume of 20 µL using 10 µL iQ™ SYBR® Green Supermix1x (BIO-RAD Laboratories; containing dNTPs, iTaq™ DNA Polymerase, MgCl₂, SYBR® Green I, enhancers, stabilizers, and passive reference dye fluorescein), 0.2 µL of each primer (10 µM), 2 µL DNA, and 7.6 µL Molecular Biology grade water (Sigma Aldrich, USA). Three technical replicates of each sample were used in all cases to improve accuracy and precision. The thermal cycler's amplification conditions were adjusted according to each gene's properties, and different annealing conditions were applied based on primers' melting temperatures.

Table S1. Primer sequences used for qPCR amplification in the study and their respective amplification sizes.

Target gene	Primer name	Sequence (5'-3')	Amplicon size (bp)	Reference
16S rRNA	338F	ACTCTACGGGAGGCAGCAG	146	(Lane, 1991)
	518R	ATTACCGCGGCTGCTGG	146	(Muyzer et al., 1993)
rpoB	rpoB-f-4	CGAACATCGGTCTGATCAACTC	360	(Silkie and Nelson, 2009)
	rpoB-r-2	GTTGCATGTTTCGACCCAT	360	(Silkie and Nelson, 2009)
bla _{CTXM}	CTXM-FW	CTATGGCACCACCAACGATA	103	(Marti and Balcázar, 2013)
	CTXM-RV	ACGGCTTTCTGCCTTAGGTT	103	(Marti and Balcázar, 2013)
ermB	ErmB-F	AAACTTACCCGCCATACCA	139	(Knapp et al., 2010)
	ErmB-R	TTTGGCGTGTTTCATTGCTT	139	(Knapp et al., 2010)
qnrS	qnrSrtF11	GACGTGCTAACTTGCCTGAT	118	(Marti and Balcázar, 2013)
	qnrSrtR11	TGGCATTGTTGGAACTTG	118	(Marti and Balcázar, 2013)
sul1	sulI-FW	CGCACCGGAAACATCGCTGCAC	163	(Pei et al., 2006)
	sulI-RV	TGAAGTTCGCGCAAGGCTCG	163	(Pei et al., 2006)
sul2	sulII-FW	TCCGGTGGAGGCCGGTATCTGG	191	(Pei et al., 2006)
	sulII-RV	CGGGAATGCCATCTGCCTTGAG	191	(Pei et al., 2006)
tetO	tetO-Fw	ACGGARAGTTTATTGTATACC	171	(Aminov et al., 2001)
	tetO-Rv	TGGCGTATCTATAATGTTGAC	171	(Aminov et al., 2001)
int1	intILC5-fw	GATCGGTGCAATGCGTGT	196	(Barraud et al., 2010)
	intILC1-rv	GCCTTGATGTTACCCGAGAG	196	(Barraud et al., 2010)

Table S2. qPCR thermal cycler conditions. The temperature and duration of the annealing phase were different for each gene: *qnrS*, *ermB*, *sul1*, *bla_{CTXM}* – 60°C, 30 sec; *int1*, *tetO* – 60°C, 1 min; *sul2* – 61°C, 1 min; *16S rRNA* – 55°C, 20 sec; *rpoB* – 55°C, 30 sec.

Genes	Step	T (°C)	Time	Genes	Step	T (°C)	Time
<i>qnrS</i> , <i>sul2</i> , <i>ermB</i> , <i>bla</i> _{CTXM}	1	95	5 min	<i>16S rRNA</i> , <i>rpoB</i> , <i>tetO</i> , <i>sul1</i> , <i>int1</i>	1	95	10 min
	2	95	15 sec		2	95	15 sec
	3	Mentioned in description			3	Mentioned in description	
	4	72	30 sec		4	72	10 sec
	5	80	5 sec		5	80	1 sec
	6	80	1 sec		6	95	15 sec
	7	12	59 sec		7	60	1 min
	8	12	1 sec		8	95	15 sec

Table S3. Synthetic DNA fragments (gBlocks) used for generating standard curves.

Gene	Sequence
<i>16S rRNA</i>	ACTCTACGGGAGGCAGCAGTGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCATGCCGCGTGTATGAAG AAGGCCTTCGGGTGTAAAGTACTTTCAGCGGGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCC GCAGAAGAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAAT
<i>rpoB</i>	CGAACATCGGTCTGATCAACTCCGGCATGTCACGTGCGCGTTAGAGTTAGGGAGCTTTCATAAGGAGCATTACGGCT GGTCGTGTGGTCGAGAACAAGTTGGACAAGTTAGACATCTGCCTTCCCCAGCCTGGTGAATCGTTACGCAGCAAGC ATAGTCACGAAGCGACAAATGATCCCTCTGCTGGGAGAGGCTCCGTAAGTGTGCAATCCAGACTTAGGCAAGG TTGCACGAGAGAGTTCTCTTCCAAGCTTCCCTTGCGTTGACGTGATAGCTCCAAGATGAACATGAAATCCATTCTTT TCCATCGTAGCTCGGAATTAGAGCTTAGCATGGGTGCGAACATGCAAC
<i>bla_{CTXM}</i>	CTATGGCACCACCAACGATATCGCGGTGATCTGGCCAAAGATCGTGCGCCGCTGATTCTGGTCACTTACTTCACCCA GCCTCAACCTAAGGCAGAAAGCCGT
<i>ermB</i>	AAACTTACCCGCCATACCACAGATGTTCCAGATAAATATTGGAAGCTATATACGTACTTTGTTTCAAATGGGTCAAT CGAGAATATCGTCAACTGTTTACTAAAAATCAGTTTCATCAAGCAATGAAACACGCCAAA
<i>qnrS</i>	GACGTGCTAACTTGCCTGATACGACATTCGTCAGTCAAGTTCATTGAACAGGGTGATATCGAAGGCTGCCACTTTG ATGTCGCAGATCTTCGTGATGCAAGTTTCCAACAATGCCA
<i>sul1</i>	CGCACCGGAAACATCGCTGCACGTGCTGTCGAACCTTCAAAGCTGAAGTCGGCGTTGGGGCTTCGCTATTGGTCTC GGTGTGCGGAAATCCTTCTTGGGCGCCACCGTTGGCCTTCTGTAAAGGATCTGGGTCCAGCGAGCCTTGCGGCGG AACTTCA
<i>sul2</i>	TGGAGGCCGGTATCTGGCGCCAGACGCAGCCATTGCGCAGGCGCGTAAGCTGATGGCCGAGGGGGCAGATGTGATC GACCTCGGTCCGGCATCCAGCAATCCCGACGCCGCGCTGTTTCGTCCGACACAGAAATCGCGCGTATCGCGCCGGT GCTGGACGCGCTCAAGGCAGATGGCATTCCCG
<i>tetO</i>	ACGGAGAGTTTATTGTATACAGTGGTGCAATTGCAGAACAGGGAGCGTAGATAAAGGCACAACAAGGACAGATA CAATGAATTTGGAGCGTCAAAGGGGAATCACTATCCAGACAGCAGTGACATCTTTTCAAGTGGGAGGATGTAAAAGTC AACATTATAGATACGCCATT
<i>int1</i>	GCCTTGATGTTACCCGAGAGCTTGGCACCCAGCCTGCGCGAGCAGCTGTGCGTGACGGGCATGGTGGCTGAAGG ACCAGGCCGAGGGCCGAGCGCGCTTGCGCTTCCGACGCCCTTGAGCGGAAGTATCCGCGCGCCGGGCATTCTCTG GCCGTGGTTCTGGGTTTTTGCGCAGCACACGCATTGACCGATC

Quantitative PCR statistical results

Table S4. P-values indicating the statistical significance of the difference between the genes' absolute concentration in the intracellular and the extracellular DNA from influent and the effluent samples.

Gene	Influent	Effluent
<i>16S rRNA</i>	0.005	0.0294
<i>bla_{CTXM}</i>	n/a	0.1939
<i>ermB</i>	0.003	0.02107
<i>qnrS</i>	0.005	0.3768
<i>sul1</i>	0.005	0.03038
<i>sul2</i>	0.013	0.03038
<i>tetO</i>	0.005	0.0294
<i>int1</i>	0.005	0.0294
<i>rpoB</i>	0.005	0.376

Table S5. P-values indicating the statistical significance of the differences between the genes' absolute concentration in the different microbial aggregates that constitute the AGS process. LG = Large granules > 1mm, SG = Small granules < 1 mm.

BEGINNING OF AERATION			
Gene	LG vs SG	BG vs flocs	SG vs flocs
<i>16S rRNA</i>	0.030	0.061	0.810
<i>bla_{CTXM}</i>	n/a	n/a	n/a
<i>ermB</i>	1.000	1.000	0.661
<i>qnrS</i>	0.312	0.030	0.194
<i>sul1</i>	0.471	0.885	0.885
<i>sul2</i>	0.665	0.030	1.000
<i>tetO</i>	0.312	0.312	0.665
<i>int1</i>	0.471	0.885	0.665
<i>rpoB</i>	0.381	0.312	0.059
END OF AERATION			
Gene	LG vs SG	BG vs flocs	SG vs flocs
<i>16S rRNA</i>	0.005	0.379	0.005
<i>bla_{CTXM}</i>	n/a	n/a	n/a
<i>ermB</i>	0.028	0.468	0.029
<i>qnrS</i>	0.005	0.005	0.575
<i>sul1</i>	0.936	0.005	0.008
<i>sul2</i>	0.066	0.045	0.936
<i>tetO</i>	0.005	0.013	0.810
<i>int1</i>	0.297	0.936	0.575
<i>rpoB</i>	0.005	0.574	0.005

Metagenomics procedure and quality control

Figure S1. Schematic representation of the procedures followed for the analysis of the metagenomics dataset.

Figure S2. Cumulative sequence counts of the raw Illumina paired-end reads showing the unique reads in blue and the duplicated reads in black.

Dimension reduction of metagenomics datasets

Figure S3. Dimension reduction of the absolute microbiome composition on genus-level after normalization and

variance stabilization transformation, with Bray-Curtis as distance metric in all methods. (A) shows a scree plot based on PCoA, showing the proportion of variance in every principal coordinate. (B) and (C) show the stress at the final iteration over ten tested dimensions for t-SNE and NMDS, respectively. (D), (E) and (F) show the dimension reductions for PCoA, t-SNE, and NMDS, respectively. Colours indicate types of samples: small granule (blue), large granule (red), floc (orange), wastewater (black), and exDNA (green). Shapes indicate stages in the WWT operation: influent (filled circle), beginning of aeration (asterisk), end of aeration (crossed square) and effluent (empty circle).

Taxonomic classification of contigs from all AGS stages and DNA fractions

Figure S4. Taxonomic classification at genus-level of short-reads for different DNA fractions and different microbial aggregate types across four stages of AGS WWT operation. A newly adapted database was used to determine relative abundance (%) of classified reads on genus-level within a sample. Labels: (F) floccular aggregates, (LG) large granules, (SG) small granules, (-1) first biological replicates, and (-2) second biological replicates. Asterisk (*) after genera names indicates genera with >3% relative abundance in at least one of the samples, also with the standard Kraken 2.0 database. Diamond (◊) after genera names indicates genera with >3% relative abundance in at least one of the samples, but not with the standard Kraken 2.0 database. Genera with <1% relative abundance in all samples are labelled as "Tax_below_1perc" (light-blue).

Taxonomic classification of ARGs contigs from all AGS stages and DNA fractions

Figure S5. Relative abundance of contigs containing ARGs across the AGS WWT operation annotated at genus-level using the newly constructed database. Bright colors (*) represent genera above 3% that were already present above 3% relative abundance in the general taxonomic classification (Figure S3). Pastel colors (◊) represent genera above 3% that were identified below 3% in the general taxonomic classification. Grey-scale colors represent genera above 3% but identified below 1% in Figure 1c. Genera with <3% relative abundance in all samples are labelled as "Tax_below_3perc" (light-blue).

1126

Reconstructed antibiotic resistance pathways from the metagenomic datasets

Figure S6. Vancomycin resistance pathway based on KEGG orthology. Colors indicate the genes that were found to be present in a certain sample: influent iDNA (blue), influent exDNA (yellow), effluent iDNA (green), and effluent exDNA (red).

1131
1132
1133
1134
1135

Figure S7. Cationic antimicrobial peptide (CAMP) resistance pathway based on KEGG orthology. Colors indicate the genes that were found to be present in a certain sample: influent iDNA (blue), influent exDNA (yellow), effluent iDNA (green), and effluent exDNA (red).

Figure S8. The beta-lactam resistance pathway based on KEGG orthology. Colors indicate the genes that were found to be present in a certain sample: influent iDNA (blue), influent exDNA (yellow), effluent iDNA (green), and effluent exDNA (red).

Quantitative PCR supplementary results

Figure S9. Relative abundance of gene copies using *16S rRNA* and *rpoB* as a proxy for the concentration of the bacteria. Genes are quantified in terms of $\log_{10}$ gene copies / $\log_{10}$ *16S rRNA* or *rpoB* gene copies. The four different graphs represent the four AGS WWT steps: (a) influent, (b) beginning of the aeration, (c) end of the aeration, and (d) effluent.

Figure S10. Ratio of intracellular over extracellular DNA concentration of the genes measured in $\log_{10}$ gene copies mL^{-1} for the influent and the effluent phase of the AGS WWT process. Ratios labelled with N/A indicate that the gene was not detected in at least one of the two DNA fractions.

Figure S11. Effluent minus influent (Δ) in $\log_{10}$ gene copies mL^{-1} for the extracellular (left) and intracellular (right) DNA fraction. The significance of the data, as calculated by the non-parametric Wilcoxon test, is displayed as $p < 0.05$ (*) and $p < 0.005$ (**).

ARGs and MGEs co-localization results

Table S6. Identified contigs containing multiple ARGs. Some also contain MGEs. Annotation of contigs by alignment in ISfinder database is provided in the “Insertion sequence” column with corresponding ISfinder nomenclature. Annotation of contigs by alignment in INTEGRAL database is provided in the “Integron” column with corresponding NCBI accession number.

Sample	Query(bp)	Influent			Beginning of aeration			End of aeration			Effluent		
		Sample	Query(bp)	ARG	IS	Integron	Sample	Query(bp)	ARG	IS	Integron	Sample	Query(bp)
iDNA-1	966	iDNA-1	1,812	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	-	<i>intI</i> JX566770	iDNA-1 LG	5,626	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> , AF024602	iDNA-1	1,034
iDNA-1	643	iDNA-1	2,542	<i>smr(B)</i> <i>smr(E)</i>	ISTu3	-	iDNA-1 LG	1,348	<i>mph(E)</i> <i>smr(E)</i>	-	-	iDNA-1	4,002
iDNA-1	566	iDNA-1	1,897	<i>msr(A)</i> <i>msr(D)</i>	-	-	iDNA-1 LG	9,566	<i>smr(B)</i> <i>smr(B)</i>	Elia1	-	iDNA-2	1,054
iDNA-1	2,470	iDNA-2	700	<i>aadA6</i> <i>ant(5'')-la</i>	TnA33	<i>intI</i> FN827339	iDNA-1 LG	1,596	<i>smr(B)</i> <i>smr(B)</i>	-	-	iDNA-2	3,649
iDNA-1	2,834	iDNA-2	1,970	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> AF024602	iDNA-2 LG	608	<i>aad(3)-lb</i> <i>aad(6)-ld</i>	TnA33	<i>intI</i> FN957811	eDNA-1	1,000
iDNA-1	1,844	iDNA-2	581	<i>aadA1</i> <i>laOXA-534</i>	-	FN99376	iDNA-2 LG	1,246	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	AF024602	eDNA-1	966
iDNA-1	751	iDNA-1	1,381	<i>aadA5</i> <i>aadA6</i>	ISP42	NC_013951	iDNA-2 LG	648	<i>mph(E)</i> <i>smr(E)</i>	-	-	eDNA-1	614
iDNA-1	1,025	iDNA-1	931	<i>aadA6</i> <i>aadA6</i>	TnA33	JQ837991	iDNA-2 LG	3,533	<i>smr(B)</i> <i>smr(B)</i>	ISV33	FJ442935	eDNA-1	2,041
iDNA-1	3,352	iDNA-1	679	<i>aadA6</i> <i>aadA6</i>	ISCw7	FN824608	iDNA-1 SG	747	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	AF024602	eDNA-1	2,819
iDNA-1	3,239	iDNA-1	1,761	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> FJ442934	iDNA-1 SG	1,612	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	TnA33	<i>intI</i> PSEADA	eDNA-1	1,043
iDNA-1	2,614	iDNA-1	3,912	<i>smr(B)</i> <i>smr(B)</i>	-	-	iDNA-1 SG	1,293	<i>aadA6</i> <i>aadA6</i>	TnA33	JQ837991	eDNA-1	2,834
iDNA-1	1,834	iDNA-1	2,347	<i>mph(E)</i> <i>smr(E)</i>	-	-	iDNA-1 SG	2,028	<i>smr(B)</i> <i>smr(B)</i>	-	-	eDNA-1	1,093
iDNA-1	1,618	iDNA-1	2,069	<i>car(A)</i> <i>smr(B)</i>	ISMycal	-	iDNA-1 SG	2,123	<i>aph(3'')-lb</i> <i>smr(B)</i>	Tn5393	AF024602	eDNA-1	3,341
iDNA-1	1,101	iDNA-1	2,543	<i>ter(A)</i> <i>su12</i>	TnA33	FJ442935	iDNA-2 SG	607	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	KJ131499	eDNA-1	968
iDNA-1	982	iDNA-2	922	<i>aadA3</i> <i>aadA4</i>	-	<i>intI</i> NC_013951	iDNA-2 SG	533	<i>mph(E)</i> <i>mph(G)</i>	-	-	eDNA-1	3,014
iDNA-2	2,927	iDNA-2	3,127	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> AF024602	iDNA-1 Flows	1,028	<i>aadA6</i> <i>aadA6</i>	TnA33	JQ837991	eDNA-1	2,594
iDNA-2	561	iDNA-2	2,127	<i>car(A)</i> <i>car(A)</i>	-	-	eDNA-1	2,929	<i>aad(3)-lb</i> <i>aadA11</i>	TnA33	<i>intI</i> JX566770	eDNA-1	681
iDNA-2	951	iDNA-2	635	<i>ter(A)</i> <i>ter(A)</i>	ISAar26	-	eDNA-1	2,191	<i>aadA6</i> <i>aadA6</i>	TnA33	-	eDNA-1	1,754
iDNA-2	4,148	iDNA-2	614	<i>aac(3)-lb</i> <i>aadA8</i>	TnA33	Y14748	eDNA-1	1,577	<i>aadA11</i> <i>aadA16</i>	TnA33	JX494726	eDNA-2	942
iDNA-2	3,012	iDNA-2	543	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	TnA33	JX494726	eDNA-1	864	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	KJ131499	eDNA-2	2,660
iDNA-2	4,392	iDNA-2	1,218	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	KJ131499	eDNA-1	5,595	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> AF024602	-	-
iDNA-2	11,139	iDNA-2	1,026	<i>aadA6</i> <i>aadA6</i>	TnA33	-	eDNA-2	3,394	<i>smr(B)</i> <i>smr(B)</i>	-	-	-	-
iDNA-2	786	iDNA-2	-	-	-	-	eDNA-2	1,833	<i>mph(E)</i> <i>mph(E)</i>	-	-	-	-
iDNA-2	658	iDNA-2	-	-	-	-	eDNA-2	2,526	<i>aph(3'')-lb</i> <i>aph(6)-ld</i>	Tn5393	<i>intI</i> AF024602	-	-
iDNA-2	1,694	iDNA-2	-	-	-	-	eDNA-2	1,646	<i>aadA10</i> <i>aadA6</i>	TnA33	JX494726	-	-
iDNA-2	3,714	iDNA-2	-	-	-	-	eDNA-2	3,706	<i>smr(B)</i> <i>smr(B)</i>	-	-	-	-
iDNA-2	786	iDNA-2	-	-	-	-	eDNA-2	2,292	<i>mph(E)</i> <i>smr(B)</i>	-	-	-	-
eDNA-1	584	eDNA-1	-	-	-	-	eDNA-2	581	<i>ter(40)</i> <i>ter(2)/W/O</i>	-	-	-	-
eDNA-1	3,217	eDNA-1	-	-	-	-	-	-	-	-	-	-	-

1163 **Table S7.** qPCR results for the selected panel of ARGs and MGE during the AGS WWT operation process.

Gene		16S rRNA				rpoB			
Sample	Biological replicate	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD
Influent exDNA	1	18.62	4.95	5.03	0.72	31.56	1.97	2.83	0.56
	2	20.95	4.20			27.58	3.05		
	3	14.70	5.94			26.13	3.45		
Influent iDNA	1	9.96	7.55	7.21	0.27	19.97	5.42	5.20	0.16
	2	11.27	7.19			21.04	5.13		
	3	12.34	6.89			21.31	5.06		
Beginning of aeration exDNA	1	18.19	4.97	4.31	0.51	31.30	2.04	2.02	0.03
	2	22.62	3.74			-	-		
	3	20.91	4.22			31.51	1.99		
Beginning of aeration granules >1mm	1	-	-	11.18	0.02	-	-	9.26	0.14
	2	14.20	11.18			23.83	9.18		
	3	14.21	11.18			23.21	9.35		
Beginning of aeration granules 0.2-1mm	1	12.51	11.59	11.21	0.27	22.51	9.48	9.19	0.28
	2	14.55	11.03			23.30	9.27		
	3	14.55	11.03			24.96	8.81		
Beginning of aeration flocs	1	13.37	11.38	11.15	0.17	23.61	9.21	9.03	0.15
	2	14.51	11.06			24.35	9.00		
	3	14.71	11.01			24.81	8.88		
End of aeration exDNA	1	19.09	4.72	4.35	0.33	-	-	1.33	0.14
	2	20.25	4.40			31.56	1.97		
	3	21.95	3.93			31.37	2.02		
End of aeration granules >1mm	1	13.54	11.40	11.25	0.19	24.16	9.12	9.18	0.07
	2	13.76	11.34			24.19	9.12		
	3	14.98	11.00			23.58	9.28		
End of aeration granules 0.2-1mm	1	13.63	9.53	9.62	0.16	23.24	7.53	7.57	0.07
	2	12.66	9.80			23.00	7.59		
	3	13.64	9.52			23.04	7.58		
End of aeration flocs	1	13.38	11.43	11.38	0.18	23.76	9.22	9.27	0.19
	2	13.60	11.37			24.18	9.11		
	3	13.71	11.34			22.77	9.49		
Effluent exDNA	1	16.79	5.36	5.05	0.30	27.56	3.07	3.55	0.48
	2	19.03	4.74			24.01	4.02		
	3	-	-			-	-		
Effluent iDNA	1	12.43	6.87	6.40	0.35	22.72	4.67	3.81	0.74
	2	14.68	6.24			30.41	2.59		
	3	15.23	6.09			24.36	4.16		

Gene		bla _{CTXM}				ermB			
Sample	Biological replicate	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD
Influent exDNA	1	-	-	-	-	-	-	-	-
	2	-	-			-	-		
	3	-	-			-	-		
Influent iDNA	1	27.97	3.79	2.98	0.59	19.69	5.22	4.73	0.71
	2	31.77	2.66			19.25	5.35		
	3	32.32	2.49			24.88	3.61		
Beginning of aeration exDNA	1	-	-	-	-	32.34	1.01	1.01	0.11
	2	-	-			-	-		
	3	-	-			-	-		
Beginning of aeration granules >1mm	1	-	-	-	-	-	-	8.62	0.09
	2	-	-			24.22	8.62		
	3	-	-			-	-		
Beginning of aeration granules 0.2-1mm	1	-	-	-	-	20.98	9.56	9.14	0.49
	2	-	-			21.32	9.46		
	3	-	-			24.77	8.40		
Beginning of aeration flocs	1	-	-	-	-	20.39	9.77	9.19	0.50
	2	-	-			21.60	9.40		
	3	-	-			24.81	8.41		
End of aeration exDNA	1	-	-	-	-	31.59	1.24	1.42	0.19
	2	-	-			30.47	1.59		
	3	-	-			-	-		
End of aeration granules >1mm	1	-	-	-	-	23.26	8.96	8.55	0.59
	2	-	-			24.39	8.61		
	3	-	-			26.12	8.07		
End of aeration granules 0.2-1mm	1	-	-	-	-	20.68	9.78	9.72	0.06
	2	-	-			21.06	9.66		
	3	-	-			-	-		
End of aeration flocs	1	-	-	-	-	22.54	9.16	9.03	0.20
	2	-	-			23.40	8.90		
	3	-	-			-	-		
Effluent exDNA	1	-	-	-	-	28.96	2.05	2.05	0.00
	2	-	-			-	-		
	3	-	-			-	-		
Effluent iDNA	1	33.51	2.14	2.14	0.00	23.95	3.90	3.88	0.11
	2	-	-			24.09	3.86		
	3	-	-			-	-		

Gene		qnrS				sul1			
Sample	Biological replicate	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD
Influent exDNA	1	31.81	0.79	1.92	0.96	24.85	2.41	2.61	1.02
	2	27.85	1.86			28.40	1.41		
	3	23.28	3.10			19.17	4.01		
Influent iDNA	1	22.07	3.73	3.60	0.27	16.44	5.08	4.90	0.23
	2	21.62	3.85			16.67	5.01		
	3	23.94	3.22			18.13	4.60		
Beginning of aeration exDNA	1	35.55	-0.22	0.21	0.57	23.36	2.81	2.76	0.24
	2	35.23	-0.13			24.69	2.46		
	3	31.06	1			22.72	3.01		
Beginning of aeration granules >1mm	1	-	-	5.55	0.25	-	-	9.57	0.06
	2	33.98	5.31			17.40	9.61		
	3	32.21	5.79			17.74	9.52		
Beginning of aeration granules 0.2-1mm	1	30.94	6.08	5.95	0.22	17.08	9.64	9.55	0.12
	2	30.84	6.10			17.75	9.46		
	3	32.42	5.68			-	-		
Beginning of aeration flocs	1	30.47	6.23	6.17	0.12	18.16	9.37	9.50	0.14
	2	30.30	6.27			17.12	9.66		
	3	31.29	6.01			17.81	9.47		
End of aeration exDNA	1	34.82	-0.02	0.27	0.28	23.28	2.85	2.91	0.29
	2	32.58	0.58			21.93	3.23		
	3	33.79	0.26			24.01	2.65		
End of aeration granules >1mm	1	33.88	5.38	5.29	0.14	17.28	9.68	9.82	0.21
	2	34.89	5.10			15.44	10.09		
	3	33.77	5.40			17.28	9.68		
End of aeration granules 0.2-1mm	1	30.85	6.22	6.31	0.09	17.21	9.73	9.77	0.17
	2	30.45	6.33			16.33	9.98		
	3	30.33	6.36			17.68	9.60		
End of aeration flocs	1	31.25	6.07	6.25	0.18	18.12	9.43	9.43	0.12
	2	29.80	6.46			17.95	9.48		
	3	30.69	6.22			18.32	9.37		
Effluent exDNA	1	26.74	2.16	1.34	0.82	23.73	2.72	2.90	0.17
	2	32.81	0.52			22.51	3.07		
	3	-	-			-	-		
Effluent iDNA	1	28.94	1.87	1.72	0.21	17.58	4.76	4.60	0.22
	2	28.81	1.91			17.71	4.72		
	3	30.68	1.40			19.17	4.31		

Gene		sul2				tetO			
Sample	Biological replicate	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD
Influent exDNA	1	30.57	1.17	2.43	1.00	27.85	1.56	1.35	0.36
	2	25.88	2.51			30.60	0.79		
	3	22.04	3.60			27.34	1.71		
Influent iDNA	1	22.05	3.90	3.83	0.18	18.68	4.43	4.39	0.13
	2	21.80	3.98			19.33	4.25		
	3	23.10	3.60			18.43	1.71		
Beginning of aeration exDNA	1	31.83	0.81	0.69	0.44	30.46	0.83	0.67	0.28
	2	34.32	0.11			32.42	0.29		
	3	30.62	1.16			30.30	0.88		
Beginning of aeration granules >1mm	1	-	-	8.15	0.07	-	-	7.49	0.19
	2	23.99	8.16			25.94	7.20		
	3	24.01	8.15			23.91	7.77		
Beginning of aeration granules 0.2-1mm	1	24.07	8.07	7.89	0.23	23.30	7.88	7.79	0.17
	2	23.57	8.22			23.13	7.93		
	3	26.54	7.37			24.49	7.55		
Beginning of aeration flocs	1	23.99	8.12	7.85	0.23	23.54	7.84	7.83	0.20
	2	24.98	7.84			22.88	8.03		
	3	25.88	7.58			24.30	7.63		
End of aeration exDNA	1	32.40	0.65	0.72	0.38	29.65	1.06	0.90	0.33
	2	30.42	1.22			29.17	1.19		
	3	33.65	0.30			31.88	0.44		
End of aeration granules >1mm	1	23.97	8.20	8.18	0.09	26.20	7.17	7.27	0.27
	2	23.69	8.22			26.71	7.03		
	3	24.17	8.14			24.60	7.62		
End of aeration granules 0.2-1mm	1	24.92	7.96	8.07	0.09	23.56	7.93	8.02	0.12
	2	24.21	8.16			22.92	8.11		
	3	24.43	8.10			23.30	8.10		
End of aeration flocs	1	25.39	7.78	7.99	0.19	24.21	7.71	7.99	0.25
	2	24.34	8.08			22.44	8.20		
	3	24.17	8.13			22.91	8.07		
Effluent exDNA	1	26.43	2.35	1.94	0.42	33.16	0.08	0.35	0.23
	2	29.35	1.52			31.65	0.61		
	3	-	-			-	-		
Effluent iDNA	1	25.68	2.87	2.75	0.16	24.55	2.79	2.81	0.16
	2	25.80	2.84			23.78	3.00		
	3	26.83	2.54			25.08	2.64		

Gene		int1							
Sample	Biological replicate	Mean ct value	Log ₁₀ gene copies	Mean Log ₁₀ gene copies	SD				
Influent exDNA	1	23.82	2.71	3.55	0.61				
	2	19.93	3.83						
	3	18.97	4.10						
Influent iDNA	1	15.89	5.29	5.15	0.13				
	2	16.25	5.19						
	3	16.98	4.98						
Beginning of aeration exDNA	1	23.50	2.80	2.36	0.37				
	2	26.43	1.96						
	3	25.11	2.34						
Beginning of aeration granules >1mm	1	-	-	9.39	0.18				
	2	18.96	9.21						
	3	17.76	9.56						
Beginning of aeration granules 0.2-1mm	1	17.39	9.61	9.56	0.28				
	2	16.86	9.76						
	3	18.44	9.30						
Beginning of aeration flocs	1	18.69	9.26	9.33	0.10				
	2	18.47	9.32						
	3	18.16	9.41						
End of aeration exDNA	1	24.17	2.61	2.50	0.18				
	2	24.31	2.57						
	3	25.17	2.32						
End of aeration granules >1mm	1	18.58	9.36	9.46	0.16				
	2	18.62	9.35						
	3	17.50	9.67						
End of aeration granules 0.2-1mm	1	18.56	9.39	9.56	0.14				
	2	17.77	9.62						
	3	17.63	9.66						
End of aeration flocs	1	19.13	9.18	9.50	0.28				
	2	17.93	9.53						
	3	17.05	9.78						
Effluent exDNA	1	24.56	2.49	2.59	0.10				
	2	23.88	2.69						
	3	-	-						
Effluent iDNA	1	18.55	4.53	4.43	0.24				
	2	18.19	4.63						
	3	19.93	4.13						

1169
1170

Table S8. Secondary metabolite gene clusters identified from recovered metagenome-assembled genomes (MAGs) from small and large granules at the end of the aeration of the AGS WWT operation process in both bioreplicates. MAGs from floccular sludge could not be recovered.

SAMPLE																
Large granules end of aeration - Replica 1																
ID and best hit MASH	antiSMASH secondary metabolite gene clusters in bacterial genomes															
NZ_CDOE01000001.1 Capnocytophaga canimorsus genome assembly	terpene	indole														
NC_014355.1 Candidatus Nitrospira defluvioli chromosome, complete genome	phosphonate	terpene 1	terpene 2	terpene 3	RIPP-like	NRPS										
NC_029602.1 Lake Sarah-associated circular molecule 3 isolate L5aCM-3-LSMU-2013, complete sequence	-															
NZ_KE386810.1 Geothrix fermentans	aryl/polyene	RIPP-like														
NC_032182.1 Hubei sobemo-like virus	-															
NZ_HG422565.1 Candidatus Microthrix parvicella RN1	ectoine	RIPP-like	T1PKS	ranthipeptide												
NC_010239.1 Radish leaf curl virus satellite DNA beta, complete genome	-															
NZ_JOEE01000001.1 Phycococcus jejuensis strain NRRL B-24460	NRPS															
NZ_FAOS01000001.1 Janthinobacterium sp. CG23_2	aryl/polyene	terpene 1	terpene 2 (geosmin)	NRPS-like	terpene 3	NRPS	terpene 4	terpene 5	RRE-containi	RRE-containi	RRE-containi	T1PKS	RIPP-like	RRE-containi	RIPP-like	terpene 6
NZ_FQYN01000017.1 Hymenobacter daecheongensis DSM 21074	terpene	NRPS														
Large granules end of aeration - Replica 2																
ID and best hit MASH	antiSMASH secondary metabolite gene clusters in bacterial genomes															
NC_025709.1 Circoviridae 4 LDM-2013, complete genome	RIPP-like	RRE-containi	terpene	terpene 2	terpene 3	NRPS-like	RRE-containi	RIPP-like	T1PKS	RRE-containing						
NZ_HG422565.1 Candidatus Microthrix parvicella	Ranthipeptide	Ectoine	RIPP-like	T1PKS												
NZ_FQYN01000017.1 Hymenobacter daecheongensis DSM 21074	T1PKS	terpene (carotenoid)														
NC_014355.1 Candidatus Nitrospira defluvioli chromosome, complete genome	RIPP-like	terpene	NRPS	terpene 2	terpene 3	phosphonate										
NC_020106.1 STL polyomavirus strain MA138, complete genome	terpene															
NZ_CM003601.1 Streptomyces anulatus strain ATCC 11523	aryl/polyene	aryl/polyene	aryl/polyene 3	T3PKS	terpene	aryl/polyene 4										
NZ_AMPG01000001.1 Candidatus Microthrix parvicella Bio17-1	terpene (isorenieratene)															
NZ_CP016447.1 Acidovorax sp. RAC01, complete genome	ectoine	terpene	terpene 2	aryl/polyene	RIPP-like	resorcinol										
Small granules end of aeration - Replica 1																
ID and best hit MASH	antiSMASH secondary metabolite gene clusters in bacterial genomes															
NZ_BBUX01000034.1 Comamonas granuli NBRC 101663 DNA	RIPP-like	terpene 1	terpene 2													
NZ_FAOS01000001.1 Janthinobacterium sp. CG23_2	terpene 1	NRPS	terpene 2	terpene 3	terpene 4	RRE-containi	terpene 5	RRE-containi	RIPP-like	terpene 6	NRPS-like	RRE-containi	RIPP-like 2	RRE-containing 3		
NZ_LZPL01000001.1 Methanobrevibacter sp. A27	aryl/polyene	T3PKS	aryl/polyene 2	aryl/polyene 3	terpene											
NZ_CP010777.1 Rulibacter sp. DG310, complete genome	terpene	NRPS														
NZ_AMPG01000001.1 Candidatus Microthrix parvicella Bio17-1	ectoine	ranthipeptidi	RIPP-like	T1PKS	T1PKS 2											
NC_014355.1 Candidatus Nitrospira defluvioli chromosome, complete genome	terpene	terpene 2	RIPP-like	terpene 3	NRPS											
Small granules end of aeration - Replica 2																
ID and best hit MASH	antiSMASH secondary metabolite gene clusters in bacterial genomes															
NC_010239.1 Radish leaf curl virus satellite DNA beta, complete genome	aryl/polyene	T3PKS	aryl/polyene 2	resorcinol	aryl/polyene											
NC_014355.1 Candidatus Nitrospira defluvioli chromosome, complete genome	terpene	terpene 2	phosphonate	RIPP-like	terpene 3	NRPS-like	NRPS									
NZ_FAOS01000001.1 Janthinobacterium sp. CG23_2	terpene	NRPS	NRPS-like	aryl/polyene	NRPS-like	aryl/polyene	terpene 2	terpene 3	RRE-containi	terpene 4	RRE-containi	terpene 5	RRE-containi	T1PKS	terpene 5	RRE-containi
NZ_AMPG01000001.1 Candidatus Microthrix parvicella Bio17-1	RRE-containing	T1PKS	T1PKS	RIPP-like												

1171
1172
1173
1174
1175